

AD-785 600

PROCEEDINGS OF THE 1974 ARMY SCIENCE
CONFERENCE HELD AT UNITED STATES
MILITARY ACADEMY, WEST POINT, NEW YORK,
18-21 JUNE 1974. VOLUME I. PRINCIPAL
AUTHORS A THROUGH H

Office, Chief of Research and Development
Washington, D. C.

26 August 1974

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE
5285 Port Royal Road, Springfield Va. 22151

AD-785 600



Army Science Conference Proceedings

18 - 21 June 1974

Volume I

Principal Authors A thru H



This document has been approved for public
release and sale; its distribution is unlimited.

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151



DEPARTMENT OF THE ARMY
OFFICE OF THE ADJUTANT GENERAL
WASHINGTON, D.C. 2031J

IN REPLY REFER TO

DAAG-PAP-A (M) (30 Jun 74) DAMA-ARZ

26 August 1974

Expires 26 August 1975

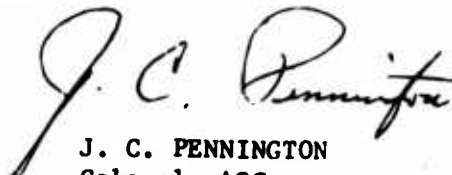
SUBJECT: 1974 Army Science Conference, Volume I

SEE DISTRIBUTION

1. Inclosed for your information and use is Volume I of the 1974 Army Science Conference Proceedings. This volume contains the unclassified papers by principal authors A thru H which were presented at the conference, 18-21 June 1974, U. S. Military Academy, West Point, New York.
2. Volumes II, III and IV of the Proceedings are being distributed separately.
3. Chiefs of Army Staff Agencies, major commanders, and heads of laboratories on the distribution list will insure that copies of the documents are placed in Technical Libraries where they will be available for reference.

BY ORDER OF THE SECRETARY OF THE ARMY:

1 Incl
Vol I, 1974 Army Science
Conference Proceedings


J. C. PENNINGTON
Colonel, AGC
Acting The Adjutant General

PROCEEDINGS
OF THE
1974 ARMY SCIENCE CONFERENCE
UNITED STATES MILITARY ACADEMY, WEST POINT, N.Y.

18 - 21 JUNE 1974

VOLUME I

Principal Authors A through H

ib

INCL

All experiments involving live animals that are reported in the Proceedings were performed in accordance with the principles of laboratory animal care as promulgated by the National Society of Medical Research.

ic

TABLE OF CONTENTS
PROCEEDINGS OF THE 1974 ARMY SCIENCE CONFERENCE

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Adams, W. A.	See Joebstl, J. A.	2	59
Aitken, G. W. Waterhouse, R. W. Fellers, G. E.	An Investigation of the Stability of Mortars on Soil Foundations	4	1
Akers, W. A.	See Spencer, T. S.	3	147
Alesi, A. L. Ames, R. P. Gagne, R. A. Litman, A. M. Prifti, J. J.	New Materials and Construction for Improved Helmets	1	1
Allen, F. J. Braerman, W. F. Stumpf, C. R.	Initiation and Propagation of Laser Supported Combustion Waves	1	17
Altheide, C. P.	See Marvin, E. L.	2	281
Ames, R. P.	See Alesi, A. L.	1	1
Aronoff, A. D. Boghossian, W. H. Jenkinson, H. A.	electrostatic Means for Intrusion Detection and Ranging	1	33
Ballato, A. Lukaszczak, T.	A New Monolithic Crystal Fil- ter Using Piezoelectric Coupling Between Anisotropic Laminae	1	49
Ballhaus, W. F. Caradonna, F. X.	Numerical Simulation of Tran- sonic Flow About Airplanes and Helicopter Rotors	1	65
Barber, V. C.	See Murphy, N. R., Jr.	2	347
Barr, D. B.	See Illinger, J. L.	2	1
Bassett, H. L.	See Smith, B. L.	3	129
Bayles, S. F.	See Spencer, T. S.	3	147

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Best, P. R., Jr.	See Isgrig, F. A.	2	17
Bingham, G. J.	An Analytical Evaluation of Airfoil Sections for Helicopter Rotor Applications	1	81
Blais, M.	See Downs, D. S.	1	277
Bloom, K.	See Zwick, H	3	425
Boghossian, W. H.	See Aronoff, A. D.	1	33
Bond, J. W., Jr.	Hypervelocity Impact Shock- Induced Damage to Steel Armor	1	97
Bossle, P. C. Dudley, G. E. Lennox, W. J. Miller, J. I. Sultan, W. E.	Possible Incapacitants for Limited Operations	4	17
Bradshaw, R. J., Jr.	See Huang, C. C.	1	463
Braerman, W. F.	See Allen, F. J.	1	17
Brody, P. S. Crowne, F.	Large Polarization Dependent Voltages in Ferroelectric Ceramics	1	111
Bromborsky, A.	See Genuario, R. D.	4	73
Brownell, A. S. Stuck, B. E.	Ocular and Skin Hazards from CO ₂ Laser Radiation	1	123
Bucci, T. J.	See Morrissey, R. L.	2	307
Bulusu, S.	See Sharma, J.	3	69
Burchfiel, J. L.	See Sim, V. M.	3	85
Burden, H. S.	Vaporization Waves in Impulsively Heated Materials	1	139
Buser, R. G. Rohde, R. S.	Dynamic Effects in Thermal Blooming in High Power Laser Propagation	1	155
Bushnell, M. J.	See Salomon, L. L.	3	27

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Campagnuolo, C. J.	See Leupold, H. A.	2	17
Caradonna, F. X.	See Ballhaus, W. F.	1	65
Choi, C. S.	See Prask, H. J.	2	425
Chrepta, M. M.	See Jacobs, H.	2	31
Chu, S.	A More Rational Approach to the Stress Analysis of Gun Tubes	1	171
Ciurej, T. F.	See Sacco, W. J.	3	1
Clayton, B. J.	Fluidic Directional Control HONEST JOHN	4	29
Colwell, G. T.	See Smith, B. L.	3	129
Costain, T.	See Voreck, W. E., Jr.	3	259
Covington, T. G. Wegner, A. H.	A Methodology/instrumentation Concept for Total System Test and Evaluation of Army Developmental Materiel	1	187
Cowley, R. A.	See Sacco, W. J.	3	1
Crow, L. H.	Tracking Reliability Growth	1	203
Crowne, F.	See Brody, P. S.	1	111
Dalrymple, E. W.	See Voreck, W. E., Jr.	3	259
Deitz, P. H.	Image Formation by Means of Spatial Intensity Correlations	1	217
Desai, C. S. Sandhu, R. S.	Finite Element Analysis of Un- confined Flow with a Variable Time-Step Procedure	1	233
Dietrich, A. M. Greenhut, V. Golaski, S.	A Method of Controlling Shock Induced Damage Aluminum Alloys	1	247
Dietz, J. F. W. Merkel, G.	Simulation of Simple Lorentz Plasma with a Random Distribu- tion of Inductively-Loaded Dipoles	1	261

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Doali, J. O.	See Juhasz, A. A.	2	103
Doherty, R. W.	See Pinto, J. J.	4	153
Downs, D. S. Garrett, W. Wiegand, D. A. Gora, T. Blais, M. Forsyth, A. C. Fair, H. D., Jr.	Photo- and Electric Field Effects in Energetic Materials	1	277
Dudley, G. E.	See Bossle, P. C.	4	17
Duffy, F. H.	See Sim, V. M.	3	85
Eaton, J. C., Jr.	See Hegyeli, A. F.	1	431
Eddy, G. A. Kastello, M. D. Scott, S. K. Terrell, T. G.	Bolivian Hemorrhagic Fever in Rhesus Monkeys; Treatment with Specific, Homologous Antibody	1	293
Eldridge, B. F.	See Le Duc, J. W.	2	193
Empson, R. N.	See Morrissey, R. L.	2	307
Epstein, A. S.	See Polimadei, R. A.	2	395
Fair, H. D., Jr.	See Downs, D. S.	1	277
Fasig, J. W.	Automatic Shell Fragment Measuring Technique	1	303
Fellers, G. E.	See Aitken, G. W.	4	1
Fine, J. E.	See Leupold, H. A.	2	217
Fishbein, W. Foiani, D. L. Taylor, P. B.	Multisensor Mortar Locating System	4	44
Fishburn, B.	See Slagg, N.	3	113
Foiani, D. L.	See Fishbein, W.	4	44
Forsyth, A. C.	See Downs, D. S.	1	277

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Frederick, D. L. Higgins, J. E.	Rapid Case Design Using Finite Element Stress Analysis and NASTRAN	1	315
Frishman, F.	On the Arithmetic Means and Variances of Products and Ratios of Random Variables	1	331
Gabel, M. L.	See Spencer, T. S.	3	147
Gage, H. M.	See Wall, W. A.	3	275
Gagne, R. A.	See Alesi, A. L.	1	1
Gaon, M. D.	See Sim, V. M.	3	85
Garrett, W.	See Downs, D. S.	1	277
Gates, R. H. Leahy, E. J. Reed, H. H. Harvey, W. T.	Nuclear Cratering Device Simulation	4	59
Genuario, R. D. Bromborsky, A. Rosado, J. A. Tomkins, J. E.	Internal Electromagnetic Pulse Investigations at the Aurora Flash X-Ray Facility	4	73
Gilbert, E. E.	TNT By-Product Isomer Recovery for Credit	1	347
Gloeckler, F. M., Jr.	Recent Translocation Results Using Navigation Satellites	1	357
Golaski, S.	See Dietrich, A. M.	1	247
Goldfarb, M. A.	See Sacco, W. J.	3	1
Goldman, R. F.	Environment, Clothing and Personal Equipment, and Military Operations	1	371
Gonano, J. R.	Concealed Explosives Detection by Means of Nuclear Resonances	4	89
Gora, T.	See Downs, D. S.	1	277
Gora, T.	See Sharma, J.	3	69

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Gorman, F. J., Jr.	See Soicher, H.	3	139
Gowins, G. E.	See Polimadei, R. A.	2	395
Grant, J. W.	A Technique for the Validation of Vehicle Models Using the Road Simulator	1	385
Greenhut, V. A.	See Dietrich, A. M.	1	247
Halpin, S. M.	See Johnson, E. M.	2	73
Hammond, C. E.	Helicopter Ground Resonance Analysis in Light of Army Requirements	1	399
Hardin, C. S. Imeson, T. C.	Detection and Identification of Trace Quantities of Organic Vapors in the Atmosphere by Ion Cluster Mass Spectrometry and the Ionization Detector System	1	415
Hartman, R. L.	See Wilkinson, E. L.	3	331
Harvey, W. T.	See Gates, R. H.	4	59
Hegyeli, A. F. Eaton, J. C., Jr.	Cell and Organ Culture Systems to Evaluate Biodegradation and Biocompatibility of Polymers	1	431
Hibler, W. D., III	A Sea Ice Terrain and Mobility Model	1	447
Higgins, J. E.	See Frederick, D. L.	1	315
Holst, G. C.	See Zwick, H.	3	425
Huang, C. C. Bradshaw, R. J., Jr.	Hydraulic Transient Pressures in Piping	1	463
Illinger, J. L. Lewis, R. W. Barr, D. B.	Effect of Interlayer on the Impact Resistance of Acrylic/Polycarbonate Laminates	2	1
Imeson, T. C.	See Harden, C. S.	1	415
Iqbal, Z.	See Prask, H. J.	2	425

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Isgrig, F. A. Best, P. R., Jr.	Night Nap-of-the-Earth Flight Training	2	17
Jacobs, H. Chrepta, M. M.	Dielectric Waveguides for Millimeter-Wave Integrated Circuits	2	31
Jasper, L. J., Jr.	Radial Beam Microwave Amplifier	2	45
Jenkinson, H. A.	See Aronoff, A. D.	1	33
Joebstl, J. A. Walker, G. W. Adams, W. A.	Investigation of Electrocatalysts by Low Energy Electron Diffrac- tion and Related Techniques	2	59
Johnson, E. M. Halpin, S. M.	Multistage Inference Models for Intelligence Analysis	2	73
Johnston, J. V.	Compliant Surfaces for Air Bearing Gyros	2	87
Juhasz, A. A. Doali, J. O. Rocchio, J. J.	Propellant Analyses, New Approaches, Improved Cap- abilities	2	103
Junkin, W. P. King, P. V.	Suppressive Shielding of Hazardous Ammunition Operations	2	119
Kastello, M. D.	See Eddy, G. A.	1	293
Kennedy, T. E. Walker, R. E.	Experimental Determination of Shock and Vibration Levels Inside the Safeguard PARB	2	135
King, P. V.	See Junkin, W. P.	2	119
Kleider, A.	Wire Obstacle Detection Technique for Rotary Wing Aircraft	2	149
Kolobielski, M.	New Methods for Determining Additives and Contaminants in Petroleum Fuels	2	165
Laing, E. J.	Determination of Aircraft Cabin Radiation, Conduction, and Convection Heat Transfer Coefficients	2	179

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Lampi, R.	See Rowley, D.	2	531
Leahy, E. J.	See Gates, R. H.	4	59
Le Duc, J. W. Suyemoto, W. Eldridge, B. F. Russell, P. K.	Transovarial Transmission of California Encephalitis by Floodwater Mosquitoes	2	193
Lee, H.	See Leupold, H. A.	2	217
Lee, R. A. Lins, W. F.	HUMAN Vibration Measuring Instrument	2	201
Lennox, W. J.	See Bossle, P. C.	4	17
Leupold, H. A. Rotwarf, F. Campagnuolo, C. J. Fine, J. E. Lee, H.	Design of Magnetic Sensors for Obtaining an Environmental Signature	2	217
Lewis, R. W.	See Illinger, J. L.	2	1
Lins, W. F.	See Lee, R. A.	2	201
Litman, A. M.	See Alesi, A. L.	1	1
Long, J. M.	See Wilmore, D. W.	3	363
Lu, P.	See Slagg, N.	3	113
Lufkin, E. G.	See Morrissey, R. L.	2	307
Lukaszek, T.	See Ballato, A.	1	49
Lundien, J. R.	Mathematical Simulation of Rayleigh Wave Generation and Propagation in Ground Media	2	233
Luzzio, A. J.	Isolation and Purification of Hepatitis Associated Antigen by Electrochromato- graphy	2	249
McDowell, E. L.	A Generalized Kinematic Hardening Theory	2	265

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
McKeel, D. W.	See Salisbury, R. E.	3	13
Mc Kyton, R. A.	See Walls, J. J., Jr.	3	291
Marvin, E. L. Altheide, C. P.	Numerical Solution of Three- Dimensional Elasticity	2	281
Mason, A. D., Jr.	See Salisbury, R. E.	3	13
Mason, A. D., Jr.	See Wilmore, D. W.	3	363
Matton, R. W.	See Wilde, A. F.	3	315
Merkel, G.	See Dietz, J. F. W.	1	261
Mikelson, D.	See Rowley, D.	2	531
Miller, J. I.	See Bossle, P. C.	4	17
Miller, J. J.	See Throop, J. F.	3	229
Miller, J. W., Jr.	New and Improved Technologies for Military Time Fuzzing Applications	2	293
Morrissey, R. L. Bucci, T. J. empson, R. N. Lufkin, E. G.	Cellular Localization of Calcium-Binding Protein	2	307
Moskowitz, H. R.	A System for Evaluation of Military Menus	2	323
Murphy, A. L., Jr.	Uniform Wind Trajectories of a Gliding Parachute Using Azimuth Homing	2	335
Murphy, N. R., Jr. Barber, V. C.	A Vehicle-Road Compatibility Analysis and Modification System (VRCAMS)	2	347
Oertel, F. H., Jr.	Sequence Laser Interferometry in a Muzzle Jet Flow Simulator	2	363
Parks, J. G.	Multi-Signature Vehicle Discrimination	4	105

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Pearce, R. H.	A Technique for Determining a Universal Drag Function for Use in Weapon Location Radars	4	121
Pearson, J. G. Pinkham, C. F. A. Redus, K. S.	Measures of Biotic Similarity and Their Applications to Clustering Techniques	2	379
Pimm, R. S.	BMD Engagement Simulation at the ABMDA Research Center	4	137
Pinkham, C. F. A.	See Pearson, J. G.	2	379
Pinto, J. J. Stuebing, E. W. Doherty, R. W. Verderame, F. D.	An Investigation of Infrared Extinguishing Aerosols for Camouflage Applications	4	153
Polimadei, R. A. Share, S. Epstein, A. S. Gowins, G. E.	Radiation Susceptibility of High Power GaAs Infrared Light- Emitting Diodes	2	395
Pollehn, H. K.	Proximity Focused Second Generation Image Intensifiers	2	409
Prask, H. J. Choi, C. S. Iqbal, Z. Trevino, S. F.	Correlation of the Vibrations of Molecules and Stability in Energetic Materials	2	425
Previte, J.	See Rowley, D.	2	531
Pries, T. H. Young, E. T. Smith, J.	Optical Techniques for the Measurement of Crosswinds	2	441
Prifti, J. J.	See Alesi, A. L.	1	1
Pruitt, B. A., Jr.	See Salisbury, R. E.	3	13
Pruitt, B. A., Jr.	See Wilmore, D. W.	3	363
Redling, T. J. Scott, L.	Ballistic Tailoring of Cased/ Consolidated Fully Telescoped Ammunitions	2	455

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Redman, C. M.	Thermal Image Projector/ Recorder	2	469
Redus, K. S.	See Pearson, J. G.	2	379
Reed, H. H.	See Gates, R. H.	4	59
Roberts, G. E.	See Tulis, M. A.	3	245
Roberts, T. G.	Mixing Gas Dynamic Lasers	2	485
Rocchio, J. J.	See Juhasz, A. A.	2	103
Rogers, J. M.	See Wilde, A. F.	3	315
Rohde, R. S.	See Ruser, R. G.	1	155
Root, R. T. Word, L. E.	Development of a Test Bed for Evaluation of Small Unit Doctrinal Alternatives in the Combat Arms	2	499
Rosado, J. A.	See Genuario, R. D.	4	73
Rosati, V. J.	20 PPS Holmium: Yttrium Lithium Fluoride Laser Transmitter	2	515
Rothwarf, F.	See Leupold, H. A.	2	217
Rowley, D. Previte, J. Wells, R. Lampi, R. Mikelson, D.	New Quality Control Techniques for the Rapid Detection of Foodborne Microorganisms	2	531
Roylance, D. K.	Shock Hardness of Graphite/ Epoxy Composites	2	545
Ruppe, D. K.	See Sumrall, G. E.	3	213
Russell, P. K.	See Le Duc, J. W.	2	193
Sacco, W. J. Goldfarb, M. A. Weinstein, M. A. Ciurej, T. F. Cowley, R. A.	Two Prognostic Indices for Trauma Patient	3	1

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Salisbury, R. E. McKeel, D. W. Mason, A. D., Jr. Pruitt, B. A., Jr. Wade, C. W. R.	Artificial Tendons for War Injuries: Construction and Tissue Response	3	13
Salomon, L. L. Trethewey, J. D. Bushnell, M. J.	Evaluation of Clouds of Airborne Fibers	3	27
Gandhu, R. S.	See Desai, C. S.	1	233
Satterfield, D. E.	Resource Allocation and Scheduling in Ballistic Missile Defense Adaptive Control Systems	3	41
Schuchardt, J. M.	See Smith, B. L.	3	129
Scott, L.	See Redling, T. J.	2	455
Scott, S. K.	See Eddy, G. A.	1	293
Scott, W. E.	Large Amplitude Motion of a Liquid Filled Gyroscope: Non- Linear Interaction of Inertial and Rossby Waves?	3	53
Share, S.	See Polimadei, R. A.	2	395
Sharma, J. Gora, T. Bulusu, S. Wiegand, D. A.	X-Ray Photoelectron Spec- troscopy of Explosive Solids	3	69
Shimmin, R. K.	See Spencer, T. S.	3	147
Sim, V. M. Duffy, F. H. Rurchfiel, J. L. Geon, M. D.	Nerve Agents and Pesticides: Value of Computer Analysis of Electroencephalograms in the Diagnosis of Exposure to Organophosphates and Chlorinated Hydrocarbons	3	85
Simon, J.	The Effect of Explosive Detonation Characteristics on Shaped Charge Performance	3	99

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Skreen, R. W.	See Wilmore, D. W.	3	363
Slagg, N. Lu, P. Fishburn, B.	Experimental and Theoretical Modeling of Fuel-Air Detonations	3	113
Smith, B. L. Bassett, K. J. Schuchardt, J. M. Colwell, G. T.	A Microwave Transparent Method of Cooling Micro- wave Components, with Practical Results	3	129
Smith, J.	See Pries, T. H.	2	441
Soicher, H. Gorman, F. J., Jr.	Ionospheric Effects During the Partial Solar Eclipse of 10 July 1972	3	139
Spencer, T. S. Bayles, S. F. Shimmin, R. K. Gabel, M. L. Akers, W. A.	Interactions Between Mosquito Repellents and Human Skin	3	147
Sprouse, J. F.	Vegetation Effluents, Micro- organisms and Material Degradation in the Tropics	3	161
Steeves, E. C.	The Behavior of Pressure Stabilized Structural Elements under Load	3	173
Stettler, J. D. Witriol, N. M.	Vibrational Energy Transfer in Hydrogen Halide Lasers	3	137
Stuck, B. E.	See Brownell, A. S.	1	123
Stuebing, E. W.	See Pinto, J. J.	4	153
Stumpf, C. R.	See Allen, F. J.	1	17
Sturek, W. B.	Boundary-Layer Studies on Spinning Bodies of Revolution	3	199
Sultan, W. E.	See Bossle, P. C.	4	17
Sumrall, G. E. Ruppe, D. K.	An Integrated Approach to Computer Aided Design and Fabrication of Distributed Parameter Microwave Devices	3	213

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Suyemoto, W.	See Le Duc, J. W.	2	193
Tauer, K. J.	See Weiss, R. J.	3	307
Taylor, P. B.	See Fishbein, W.	4	44
Terrell, T. G.	See Eddy, G. A.	1	293
Thomas, E. A.	Radar Warning Receivers for Army Aircraft	4	169
Throop, J. F. Miller, J. J.	Fatigue Behavior of Metal Laminates	3	229
Tompkins, J. E.	See Genuario, R. D.	4	73
Trethewey, J. D.	See Salomon, L. L.	3	27
Trevino, S. F.	See Prask, H. J.	2	425
Tulis, M. A. Whiting, L. D., III Roberts, G. E.	Controlled Fireballs - Effective Kill Mechanisms for Flammable Target	3	245
Verderame, F. D.	See Pinto, J. J.	4	153
Voreck, W. E., Jr. Costain, T. Dalrymple, E. W.	Advances in Explosive Train Technology	3	259
Wade, C. W. R.	See Salisbury, R. E.	3	13
Walker, G. W.	See Joebstl, J. A.	2	59
Walker, R. E.	See Kennedy, T. E.	2	135
Wall, W. A. Gage, H. M.	Determination of the Sensitivity and Specificity of Vapor Detection systems for Explosives, Narcotics and Related Compounds	3	275
Walls, J. J., Jr. Mc Kyton, R. A.	Computer Simulation of Fabricational Tolerances for Multilayer Optical Coatings	3	291
Waterhouse, R. W.	See Aitken, G. W.	4	1

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Wegner, A. H.	See Covington, T. G.	1	187
Weinstein, M. A.	See Sacco, W. J.	3	1
Weiss, R. J. Tauer, K. J.	High Damping Materials, Fundamental Aspects	3	307
Wells, R.	See Rowley, D.	2	531
Wentworth, S. E.	See Wilde, A. . .	3	315
Whiting, L. D., III	See Tulis, M. A.	3	245
Wiegand, D. A.	See Downs, D. S.	1	277
Wiegand, D. A.	See Sharma, J.	3	69
Wilde, A. F. Matton, R. W. Rogers, J. M. Wentworth, S. E.	The Preparation and Ballistic Evaluation of Transparent Polyurethane Block Copolymers Based on 2, 4-Toluene Diisocyanate	3	315
Wilkinson, E. L. Hartman, R. L.	A Novel Laser Radar Range	3	331
Williamson, A. N.	Automated Selection and Location of Terrain Features on the Basis of Spectral Characteristics	3	347
Wilmore, D. W. Long, J. M. Skreen, R. W. Mason, A. D., Jr. Pruitt, B. A., Jr.	Metabolic Rate, Ambient Temperature and Cate- cholamines: Interrelationships Following Thermal Injury	3	363
Witroil, N. M.	See Stettler, J. D.	3	187
Witt, F.	Deformation Textures of Shear Spun Shaped Charge Liners	3	377
Word, L. E.	See Root, R. T.	2	499
Yalamanchili, R. V. S.	Unsteady Heat Transfer Analysis for Any Ammunition, Gun, and Firing Schedule	3	393

<u>Author</u>	<u>Title</u>	<u>Vol</u>	<u>Page</u>
Young, E. T.	See Pries, T. H.	2	441
Zollinger, W. D.	Meningococcal Antigens: An Immunochemical Approach to the Development of a Meningitis Vaccine	3	409
Zwick, H. Holst, G. C. Bloom, K.	Disruption of Visual Function Associated with Laser Environments	3	425

NEW MATERIALS AND CONSTRUCTION FOR IMPROVED HELMETS

ANTHONY L. ALES, MR.; RICHARD P. AMES, MR.; ROGER A. GAGNE, MR.;
ALAN M. LITMAN, MR. & JOSEPH J. PRIFTI, PhD.
ARMY MATERIALS & MECHANICS RESEARCH CENTER
WATERTOWN, MASSACHUSETTS

Three new materials are now simultaneously available for exploitation in personnel armor. They are (1) XP, a highly oriented polypropylene film, (2) LMLD - a low modulus, low density glass fiber, and (3) Kevlar - a high modulus, high strength aromatic polyamide fiber. All three are superior to the older armor materials in protection capabilities against munition fragments. (1)

The label "new" often signifies that a material has not been completely characterized and that actual use experience in product form is limited or nonexistent. If further the new material has unusual properties, properties much different from prior materials, then its behavior may not be well understood, it may be difficult to process into desired forms and constructions, and its performance in the end product may not be as expected. Both XP and Kevlar fit into this category of new material. LMLD on the other hand, and E glass fiber (ordinary fiberglass now used in armor) do not differ in properties tremendously. Both combine with the same polyester resins to make bonded fabric armor. Processing and fabrication for both are alike.

This paper will consider only the XP and Kevlar materials and some problems posed by their "newness" and unusual characteristics in their application to helmets. The preparation of flat laminates and helmet constructions and their testing in several ways for "durability" and for fragment penetration resistance capabilities will be presented. The suitability of these materials and constructions thereof for helmets will be evaluated. Remaining problems to successful application will be identified.

PROBLEMS IN USING NEW MATERIALS FOR ARMOR

A. XP Film

XP is the first armor material in the form of a thin film. It is prepared from flat tubular polypropylene film by hot-stretching to about 12 times its original length and to a tube thickness of about 0.0015 inch. ⁽²⁾ Stretching increases the tensile strength from 4300 to about 50,000 psi in the stretch direction and practically no strength (40 psi) in the transverse direction. This highly orthotropic strength characteristic can be a problem in handling and assembling several hundred plies into armor since it splits easily. Also, it requires that alternate plies be cross-ply to be effective armor. Cross-plying is accomplished in a practical manner by the application of filament winding techniques. Eight inch wide film is wound at 45° in alternate directions and a pad obtained by slitting the wound material longitudinally. Pads as large as 9x21 feet have been made. Pads may be secured by stitching around the perimeter and in horizontal and vertical rows at intervals. Film pads are converted to a rigid laminate by the simultaneous application of heat and pressure followed by cooling under pressure. ⁽³⁾ No adhesive or external bonding agent is used between the plies.

As normally made, molded XP has a tendency to delaminate partially when severely flexed or subjected to large temperature changes. Such behavior raises doubts as to the durability of an XP helmet in the field. Yet some capability for delamination under ballistic impact is necessary for a high degree of resistance to penetration. Complete fusion into a solid block will produce poor armor. Reconciling these opposing requirements is the problem.

Two approaches were taken in achieving a stable rigid structure without undue loss in protective capability. The first sought to determine whether the problem could be ameliorated to a satisfactory degree by careful selection of the molding temperature and pressure. The second encapsulates a pad of XP film between rigid skins of resin-bonded fabric laminates molded together at the edges to achieve a rigid structure that would be durable and equivalent to XP in protective capabilities.

B. Kevlar Fiber

Kevlar aromatic polyamide at present is made as three fibers having different strength properties that go up to 400,000 psi in tensile strength and 19×10^6 psi in modulus. ⁽⁴⁾ Fabrics made from these fibers differ a little in ballistic resistance: the

better of these, Kevlar 29, is usually selected for armor applications.

Many bonding resins for Kevlar fabric armor have been examined by AMMRC and other laboratories. The consensus has selected the same phenolic/polyvinyl butyral resin used for the nylon fabric helmet liner. However, the interply bond strength using this resin is not always adequate. Molding temperature and pressure, and precuring of the prepregged (resin impregnated) fabric were briefly investigated using flat panels before helmets were molded. Peel strength was used as the criteria for selecting the precure and molding conditions.

MOLDING OF TEST LAMINATES & HELMETS

A. XP

Prior work indicated that temperature alone of the molding cycle variables could appreciably alter ballistic limits of laminated XP.⁽³⁾ With increasing molding temperatures beyond about 325°F, ballistic limit decreased. However, no tests had been run to determine the effect of molding temperature on the susceptibility to delaminate. A series of 12x12 inch test panels weighing 40 oz/sq ft were made in a mold at temperatures between 315°F and 360°F (all molded for 20 minutes at 1000 psi) by the Phillips Scientific Corporation and tested at AMMRC for ballistic limit velocities. A duplicate set was cut into 6x6 inch panels and tested for delamination by temperature and humidity cycling and by water immersion.

A second series of test panels, 6x6 inches, weighing 36 oz/sq ft were molded at AMMRC over the same temperature range between polished metal plates with 0.018 inch Teflon sheets between the XP and the metal plates to prevent adhesion between them. A thermocouple was inserted at mid-thickness in a corner to measure the molding temperature accurately inside the laminate rather than relying on a platen temperature. Molding at temperature was conducted for 30 minutes - sufficient to provide a minimum of 10 minutes at the indicated temperature for the center of the panel. Panels were cooled under pressure to an internal temperature of 100°F. The molding pressure was raised to about 2500 psi since higher pressures of this magnitude were needed for helmet molding and since partial internal delamination was observed on thermal stressing of the first series.

The AMMRC panels were translucent as shown in Figure 1 while the Phillips panels were striated, white and opaque. Both when molded at the highest temperatures (356-360°F) were yellow and hazy.

The difference in appearance between the two sets of panels is attributed to the better removal of air between film plies at the higher pressure. Excessive flow of XP material occurred at the 356°F molding temperature as indicated by the flash that squeezed out around the edges.

XP helmets were made using a M-1 helmet shape, production quality compression mold in a 300 ton press at the U.S. Army Natick Laboratories. This mold produced a helmet with a wall thickness of 0.2 inch that weighed about one pound. Although a 3 pound helmet was desired, this was the only compression mold available for demonstrating the molding process. Full press tonnage and the maximum temperature of 320°F obtainable from the steam supply was used. XP helmet preforms were put into the hot mold and pressed for 20 minutes, then cooled to room temperature under pressure. The size of the preform had to be made smaller than was indicated by helmet dimensions. The shear cut-off on the mold that has operated satisfactorily when nylon fabric helmets were molded was unable to cut off any excess XP with the consequence that the mold did not fully close. The removal of about 1 to 1 1/2 inches from the perimeter of the preform corrected this. The XP flowed to fill the mold and produced a helmet with a finished rim.

Two styles of preforms were tried, an oval shaped one and a "pinwheel" type shown in Figure 3 with either overlapping or butted "vanes". Sound helmets, that is, helmets bonded throughout, were obtained with both preform styles. One of the XP helmets is shown in Figure 2 together with helmets molded from the other materials.

B. Kevlar

Flat laminates consisting of 23 plies of Kevlar 29, 11 oz/sq yard plain weave fabric, preimpregnated with phenolic/polyvinyl butyral resin (22%) were prepared by press molding under 45 tons at 330°F for 45 minutes, followed by cooling prior to removal from the press. Test panels, 12x12 inches, for ballistic testing and 6x6 inch panels cut from 12x12 inch panels were prepared having a 36 oz/sq ft areal density.

Helmet constructions of the standard M-1 helmet shape were prepared by conventional bag molding procedures in an autoclave at 330°F and 250 psi pressure for 80 minutes. The mold was a silicone rubber coated, glass fabric reinforced epoxy mold that was fabricated using a plaster cast made of the male half of the XP helmet compression mold. The mold was made so spacer liners could be inserted to obtain desired wall thicknesses. Two Kevlar 29 constructions evolved, one of 23 plies of 11 oz/sq yd plain weave

fabric, and the other of 14 plies of 16 oz/sq yd woven roving weave fabric; both preimpregnated with phenolic/polyvinyl butyral resin (22 and 25 percent respectively). Pinwheel preforms were laid up in an overlap pattern within the mold; the mold put into a polyvinyl alcohol film bag; and the bag evacuated (shown in Figure 4) prior to curing in the autoclave. The helmet made with 23 plies required three consecutive moldings of 8, 8 and 7 plies to achieve a smooth, wrinkle free model with proper tapering along the helmet edge. However, 14 plies of woven roving fabric enabled the helmet to be molded in a single operation due to the fabric's better draping characteristics and lesser total bulk of the lay-up. This helmet and a test panel to provide specimens for a peel resistance test were molded at 350°F to obtain a greater resin cure and eliminate the strong residual odor of the 330°F moldings. Helmets of both construction weighed about 3 pounds.

Kevlar fabric and laminates were found not to cut cleanly by any of the usual means - scissors, rotary knives, band saws and circular saws.

C. GRP/XP/Kevlar Composite

Flat laminates were prepared by encapsulating 16 oz/sq ft of XP film, 6x6 inches between 4 plies, 8x10 inches, of conventional polyester prepregged woven roving glass fabric (GRP) and 6 plies, 8x10 inches of polyester impregnated Kevlar 29 satin weave fabric (5 oz/sq yd). This composite was cured in the autoclave at 260°F under 75 psi pressure for 90 minutes employing conventional vacuum bag molding techniques. Under these conditions the XP film plies would not be expected to bond together.

In a similar manner, M-1 helmet shaped composites were made in the same mold utilized for the Kevlar helmet. The core XP was tapered by progressively making the inner plies smaller and the largest ply end an inch from the helmet edge so that the GRP and Kevlar bonded together to form a rigid double wall shell sealed at the rim. Again a hand lay-up pinwheel overlap procedure for assembling the materials within the mold was employed. The helmets weighed about 3 pounds.

TESTING PROCEDURES AND RESULTS

A. Ballistic Penetration Resistance

1. Helmet Materials

The ballistic resistance of the new helmet armor materials was determined using flat panels with four projectiles of different

masses representative of the fragmenting munition threat (under 100 grains) to combat ground troops. The test projectiles were conventional Army fragment simulating projectiles conforming to Military Specification MIL-P-46593. Testing was conducted in accordance with Military Standard MIL-STD-662, Ballistic Test Method for Personnel Armor Material. The test determines ballistic limit velocities, V_{50} , denoting the projectile velocity where the probability of penetration is 0.50 and calculated by averaging out an equal number of velocities within a range of 125 ft/sec. that result in partial and complete penetrations. (Ballistic data now being acquired with munition fragments and confirming the superior resistance to penetration reported herein for these materials.)

The percent improvement in V_{50} values for the new materials over the standard M-1 helmet system (steel shell and laminated nylon fabric liner) is given in Table I. It is readily apparent that all four materials exhibit substantially higher resistance to penetration than the M-1 helmet system. Kevlar and XP are best closely followed by the GRP/XP/Kevlar composite. Their improvement over the M-1 helmet system is in the order of 30 to 80 percent.

2. Variation of V_{50} Ballistic Limit of Laminated XP With Molding Temperature

Only the intermediate caliber projectile was used. Figure 5 shows the combined data for the Phillips and AMMRC panels presented in terms of V_{50} ratios based on the V_{50} value obtained for the lowest temperature (315-316°F). This ratio starts to decline rapidly after 345°F, indicating that the ballistic penetration resistance of laminated XP is being degraded at the higher temperatures.

B. Durability (Permanence) Tests

1. Temperature Cycling

All test panels were prepared with two adjacent edges cut with a band saw to obtain unsealed edges as well as molded edges. With XP this procedure produced partial delamination along the edge and some fusing of the cut surfaces.

The panels in an environmental test chamber were cycled from room temperature to -76°F, then to 176°F and back to room temperature within 8 hours. Two such cycles were run. Panels were inspected visually for changes after each cycle.

There were no changes in the appearance of the Kevlar and GRP/XP/Kevlar composite. The Phillips XP panels molded at 315 to 340°F delaminated around the cut edges to a depth of 1/8 to 3/8 inch.

All panels molded at 315 to 350° maintained their integrity and rigidity but were whiter in appearance than originally and felt slightly spongy when squeezed indicating partial delamination. The 360°F panel warped and split partially along a curved surface of discontinuity that was visible before the test started.

AMMRC XP panels molded at 316 and 324°F turned white and split along the cut edges. The former puffed up to about 1 1/2" in thickness. The 337, 347, and 356°F molded panels showed no change.

2. Humidity Cycling

This test was conducted in similar fashion to the temperature cycling test. The first part of the cycle was at 95 percent relative humidity and 176°F; the second part was at the same temperature but without the high humidity.

The behavior of the Phillips XP specimens submitted to humidity cycling was similar to those subjected to temperature cycling. It should be noted that humidity cycling also involves temperature cycling (room temperature to 176° to room temperature). The panels molded at 315°F and 360°F split as before. All but the latter felt slightly spongy when squeezed. The AMMRC panels molded at 316, 326 and 338°F split with the frequency and extent of splitting decreasing with increasing temperature. The panels molded at 346 and 356°F did not change in temperature.

The Kevlar and GRP/XP/Kevlar composite panels did not change in appearance except that for the latter the Kevlar skin had puffed up away from the XP to a height of about 1/4 inch at the center.

3. Water Absorption

Water absorption for all the materials except the composite was conducted in accordance with procedures outlined in ASTM D570-63: Standard Method of Test for Water Absorption of Plastics. The composite specimens consisted of 6x6 inch XP encapsulated within 8x10 inch GRP and Kevlar skins. Cuts to expose unsealed edges were made outside the XP area.

The 24 hour water absorption values determined were:

XP	None
Kevlar	2.5 Percent
GRP/XP/Kevlar composite	5.8 Percent

Water could be forced out of the composite specimens at the cut edges by shaking or tapping. Inspection of all edges showed that the Kevlar plies could be peeled from one another with very little effort. However, the Kevlar ply adjacent to the GRP was firmly bonded.

4. Impact Damage Resistance

To determine the relative impact resistance of M-1 shape helmet constructions, a 4 inch diameter steel ball weighing 8 pounds was dropped from a height of 5 feet onto the crown of the helmet placed on a concrete floor. The amount of permanent deformation was measured and each helmet was closely examined for delamination or other damage.

No impact damage could be detected for either Kevlar helmet construction. The GRP/XP/Kevlar composite helmet suffered no permanent deformation but careful examination of the inside of the helmet revealed an area below the impact point, about 2 inches in diameter, that had a slightly altered appearance resulting presumably from the transient deformation of the helmet structure. Impact of the M-1 Hatfield steel helmet and nylon liner system produced a dent in the metal shell 7/16" deep and 1 7/8" in diameter. The XP helmet was also tested even though it was only 1/3 the weight of the others and it had been molded at a temperature too low for adequate bonding. Accordingly, it was not surprising that upon impact a dent 1 1/4 inches deep and 3 1/4 inches in diameter formed. The dented material had delaminated.

C. Peel Resistance

This test was conducted on Kevlar laminates only. The procedure followed was that of ASTM method D1876, Peel Resistance of Adhesives (T-Peel Test) except the bonded part of the specimens was 5 instead of 6 inches. Peel resistance of the laminates made with 11 oz/sq yd plain weave fabric and molded at 330°F and with 16 oz/sq yd woven fabric molded at 350°F was about 7 pounds per inch of width when pressed at 250 psi but only about 5 1/2 pounds at 2250 psi. Precuring the first fabric (prepregged) for 8 minutes at 235°F did not significantly change the peel strength. The peel strengths shown in Table II are considered adequate but not high for well-bonded durable laminates.

DISCUSSION OF RESULTS

A Protective Capabilities

The 64 to 79 percent increase in V_{50} ballistic limit of the Kevlar and XP laminates and of the composite over the standard M-1 helmet system when tested with the projectile of least mass is especially significant. A major threat to ground troops are detonating

munitions emitting small fragments traveling at high velocities. Kevlar and XP (or composite constructions having XP as the major component) helmets weighing 3 pounds (the same as the M-1 helmet system) are considered capable of stopping most and possibly all fragments at relatively short distances from burst. The M-1 helmet system is considered ineffective under these conditions.

B. XP Helmet

Molding of XP helmets, on the basis of experience with flat laminates, requires the selection of a molding temperature of 340 to 350°F and a pressure of at least 2200 psi to insure removal of air between film plies, effect adequate bonding, and restrict the reduction in V_{50} ballistic limit to under 10 percent of the maximum attainable. The temperatures measured must be that of the XP, not that of the mold. Vacuum-bagging should be tried before molding for greater assurance of air removal. There appears to be wide latitude in preform design for XP helmets provided preforms are undersized (but not underweight) and of minimum bulk to allow the mold to close since any excess XP cannot be cut by the mold. Although molded XP helmets will require no cutting of the rim, piercing techniques with heated penetrators may be needed to make holes for mounting helmet suspensions without incurring local delamination of the XP.

Laminated XP when properly molded will not delaminate on thermal stressing, is dimensionally stable and does not absorb water. The ability of 3 pound XP helmets to withstand low velocity impact as exemplified by the ball impact test remains to be determined.

Thermal stressing can distinguish between helmets molded under different conditions of temperature and pressure. As few as two cycles over a modest temperature range, say of 100°F above ambient, may suffice as a test. Translucency and color can also serve as visual guides to adequate molding. A yellowish tinge and good translucency indicate a satisfactory molding. Opaqueness with whiteness indicates too low molding temperatures and/or pressures. Very hazy yellowness indicates too high a temperature.

C. Kevlar Helmet

Kevlar fabric laminates using phenolic/polyvinyl butyral resin withstand thermal stressing and have low water absorption (2.5 percent in 24 hours). A molding pressure of 250 psi and temperature of 350°F are adequate. Heavy fabrics of good draping characteristics make molding of helmet shapes easier. A woven roving fabric of 15 oz/sq yd, and possibly up to 25 oz/sq yd, is satisfactory, and represents a lower cost fabric construction.

Kevlar fabrics and laminates of low resin content, 20 to 30 percent, were not able to be cut cleanly. Even helmets compression molded in matched steel dies (as they would be in production) will require some treatment of the as-molded fibrous edge to obtain a finished rim.

D. GRP/XP/Kevlar Composite Helmet

This composite construction was obviously not waterproof, as evidenced by the accumulation of water between the skins. The entry point was apparently between Kevlar plies whose adhesion was quite poor. Polyester is evidently not a suitable bonding resin for Kevlar fabrics. Phenolic/polyvinyl butyral would be a better choice but its use would require some intermediate bonding medium to effect the bond with the GRP since the phenolic resin inhibits the cure of the GRP's polyester. If water also enters through the thickness of the skins, surface plies, or all plies, having somewhat higher resin contents would reduce or prevent this.

CONCLUSIONS

1. The feasibility of utilizing the new materials, Kevlar 29 fabric, laminated with phenolic/polyvinyl butyral resin, and laminated XP polypropylene film in the development of combat helmets has been demonstrated on the basis of laboratory tests for fragment protection capabilities and for durability. Some problems of secondary importance remain; obvious solutions for all but one are available.

2. Processing parameters have been defined for molding helmets weighing 3 pounds that are durable and possess at least 90 percent of the fragment protection capabilities of the material.

a) XP helmets are best molded at $345 \pm 5^\circ\text{F}$ for a minimum of 30 minutes under a minimum pressure of 2200 psi. The specified temperature is that of the laminate and not a temperature measured somewhere in the mold metal. Procedures should be adjusted or modified to prevent the occlusion of air between film plies during molding. The making of holes in XP helmets for mounting suspension systems must be conducted with care using properly selected tools and techniques to insure that the new edges are sealed by fusion and no delamination occurs. Heated penetrators may be suitable hole-making devices.

b) There are no indications that the molding of laminated Kevlar 29 fabric helmets has any critical requirements. A molding pressure of 250 psi and a temperature of 350°F produced satis-

factory helmets. Procedures or devices for cleanly cutting the helmet rim, or for removing frayed fibers and sealing the rim, or for binding the rim to form a finished edge are needed.

c) Polyester resins are not satisfactory bonding materials for Kevlar fabric in the GRP/XP/Kevlar composite helmet. Phenolic/polyvinyl butyral resin is recommended with the use of a polyester-compatible adhesive or resin-coated scrim ply between the GRP and Kevlar laminates. Resin contents of 30 percent (greater for surface plies) are recommended to reduce water absorption.

3. Suitable types of tests for the quality of molded helmets are:

- a) V_{50} ballistic limit velocity and falling ball impact for the three constructions.
- b) Thermal stress and visual inspection of color and translucency for XP helmets.
- c) Water absorption for GRP/XP/Kevlar composite helmets.
- d) Peel strength for Kevlar 29 helmets.

4. Laminated Kevlar 29 fabric, laminated XP film, GRP/XP/Kevlar composite and LMLD offer substantially improved protection capabilities against the broad range of fragment masses inherent to anti-personnel munitions. In terms of V_{50} ballistic limit velocities, for the range of test projectile masses utilized, the improvement over the standard M-1 helmet system is 42 to 79 percent for Kevlar, 36 to 77 percent for XP, 33 to 75 percent for the GRP/XP/Kevlar composite, and 13 to 53 percent for LMLD.

5. The development of a combat helmet can be successfully completed with any of the helmet systems examined. The Kevlar and LMLD helmet developments are the least difficult, XP is next, and the GRP/XP/Kevlar composite is the most difficult. The one having the most promising combination of superior performance and a minimum of problems is a Kevlar helmet.

ACKNOWLEDGEMENT

The authors thank all who participated in the work reported on by this paper: Verne Stimpert and Douglas F. Mitchell for the molding of laminates and helmets; John Shippie, Elizabeth

ALESİ, AMES, GAGNE LITMAN & PRIFTI

Cilley, Salvatore Favuzza, and Frank J. Maloney for all the testing; Mary Delay, Elizabeth A. Sterling, and Astrid V. Gallagher for preparation of the paper; and Stanley T. Wierzbicki and Salvatore Nanfria for illustrations and photographs. Jack F. Furrer and Thomas H. Bresse of the U.S. Army Natick Laboratories made available their helmet mold and press and aided in the molding of the XP helmets.

REFERENCES

1. Mascianica, F. S., "Ballistic Technology of Lightweight Armor-1973," AMMRC TR 73-47, November 1973, Army Materials and Mechanics Research Center.
2. Kile, S. A., "Scale Up and Quality Control Procedures for Flexible Armor Material XP," AMMRC CTR 72-5, April 1972, Army Materials and Mechanics Research Center.
3. Kile, S. A., "Protection and Quality Control Procedures for Rigid Armor Material XP," AMMRC CTR 74-12, February 1974, Army Materials and Mechanics Research Center.
4. Alesi, A. L. et al, "Review of Application of PRD 49 Fiber to Composite Structures," 10 May 1973, Army Materials and Mechanics Research Center.

Table I. V_{50} Ballistic Limits of New Helmet Materials Compared to the M-1 Helmet

Material	Percent Improvement*			
	A	B	C	D
KEVLAR	79	42	62	56
XP	66	36	69	77
GRP/XP/KEVLAR	64	33	56	75
LMLD	53	15	40	13

*Percent Improvement = $\frac{V_{50}(\text{New Material}) - V_{50}(\text{M-1 Helmet})}{V_{50}(\text{M-1 Helmet})}$

NOTES: A, B, C, D indicate fragment-simulating projectiles in increasing order of mass.
 All comparisons are at an areal density of 36 oz/sq ft.
 M-1 helmet includes Hadfield steel shell and nylon fabric reinforced plastic liner.

Table II. Peel Resistance of KEVLAR Laminates (lb/in.)

	Molding Temperature (°F)	Molding Pressure	
		(250 psi)	(2250 psi)
Not Precured	330*	7.3	5.7
	350†	7.0	-
Precured‡	330*	7.6	5.5

*KEVLAR 29, 11 oz/sq yd plain weave fabric; 22% resin, phenolic/polyvinyl butyral.

†KEVLAR 29, 16 oz/sq yd woven roving fabric, 25% resin, phenolic/polyvinyl butyral.

‡For 8 minutes at 235°F.



Figure 1.
 Relative Translucency of
 XP Laminates: (left to right)
 AMMRC panel molded at 2250 psi
 for 30 minutes at 347°F;
 Phillips panels molded at
 1000 psi for 20 minutes at
 360°F and 315°F.

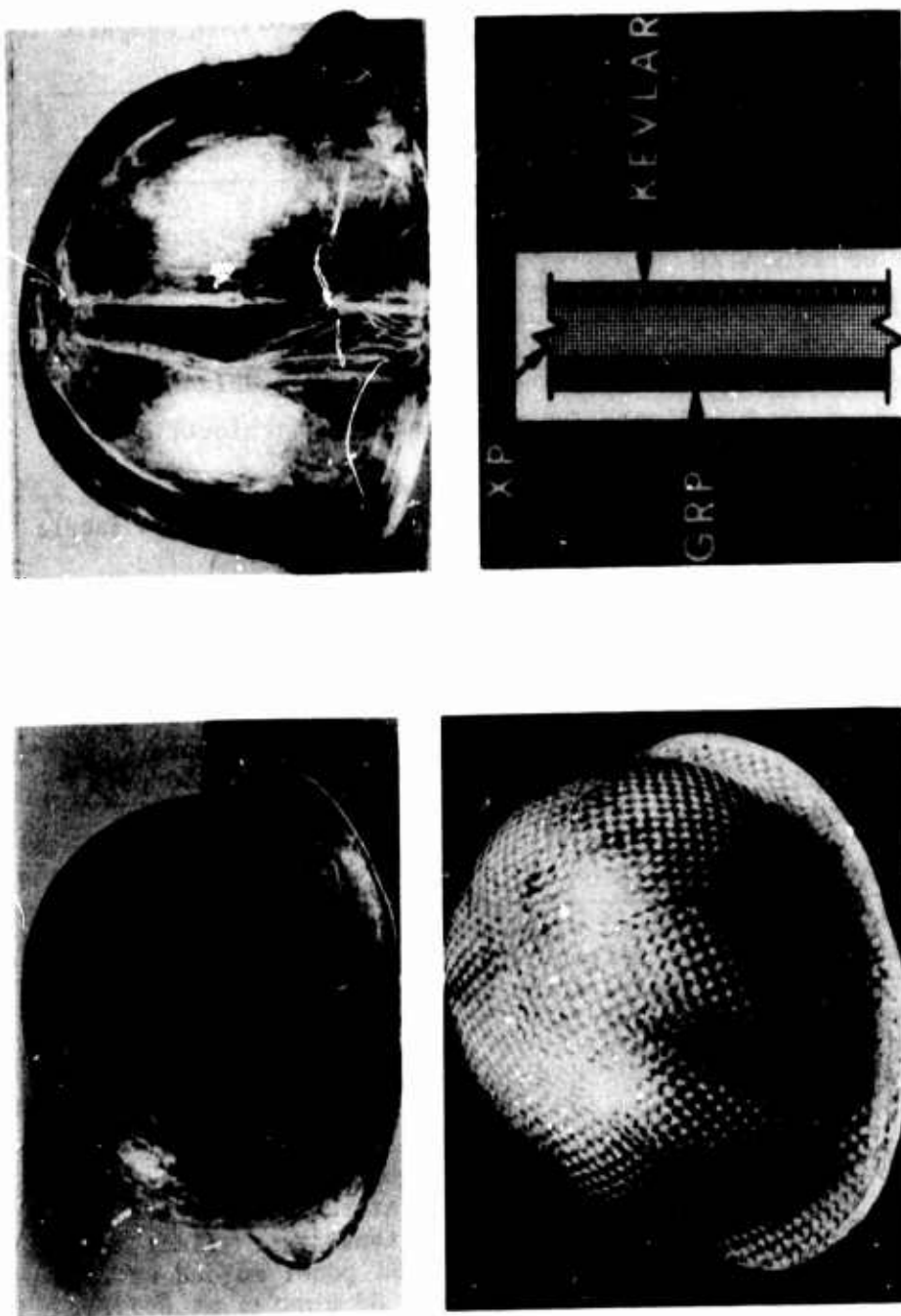


Figure 2. Helmets: KEVLAR (upper left), XP (upper right), Composite (lower left), Composite Helmet Cross Section (lower right)

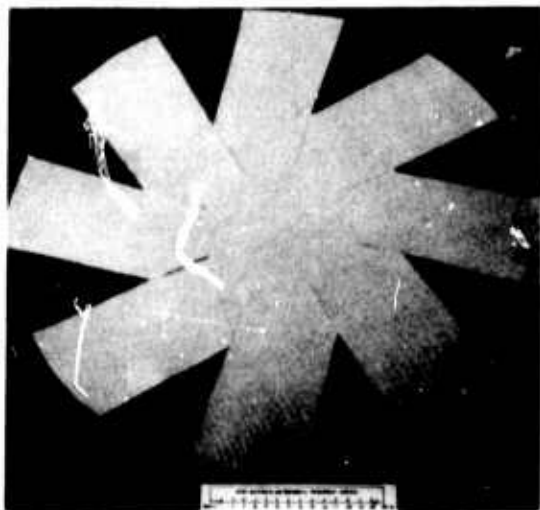


Figure 3. "Pinwheel" Preform

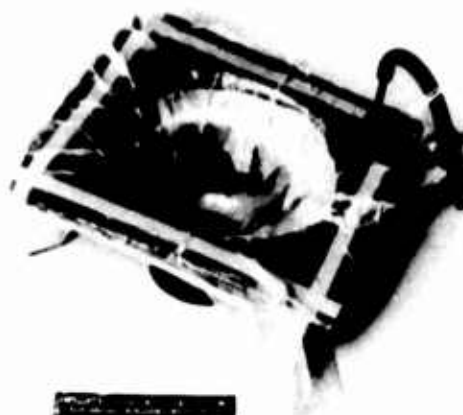


Figure 4. Vacuum-Bagged Helmet Preform in Mold

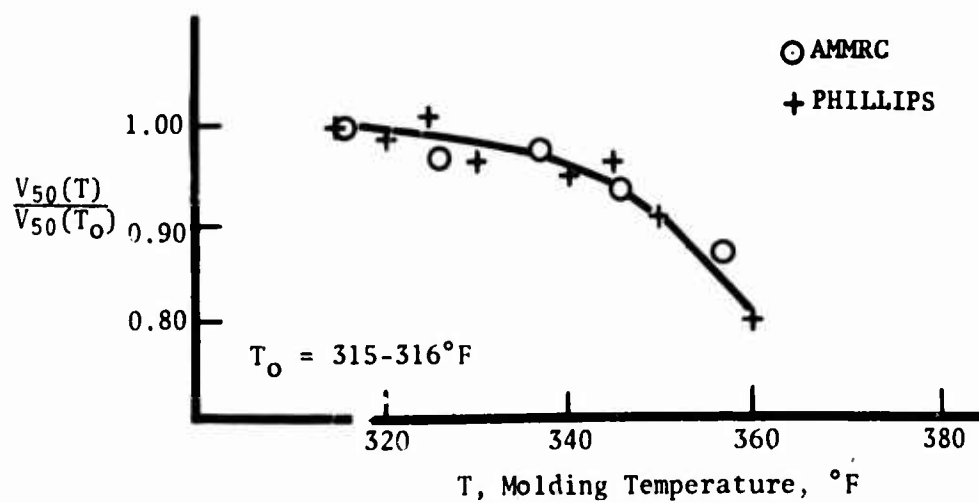


Figure 5. Effect of Molding Temperature on V₅₀ Ballistic Limit of Laminated XP

INITIATION AND PROPAGATION OF LASER SUPPORTED COMBUSTION WAVES

FRANK J. ALLEN, *WILLIAM F. BRAERMAN, CHARLES R. STUMPFEL
U. S. ARMY BALLISTIC RESEARCH LABORATORIES
ABERDEEN PROVING GROUND, MARYLAND 21005

I. INTRODUCTION

During the past few years there has been much discussion of absorption waves produced by intense laser irradiation of targets (1-6). These have been classified according to the following table (5).

Table 1. Classes of Absorption Waves

Phenomenon	Intensity (Approx.)	Propagation Velocity (Approx.)
Breakdown Wave	10^9 W/cm ²	10^6 cm/sec
Radiation Front	10^7	10^5
Laser Supported Detonation Wave	10^4	10^3
Laser Supported Combustion Wave	10^4	10^3
Plasmatron	10^4	0

In each case the absorption of laser beam energy behind the wave front supplies the energy required to overcome losses associated with propagation. In a detonation wave propagation is supersonic; in a combustion wave, subsonic. An optical plasmatron is a limiting case of a laser supported combustion (LSC) wave in which the laser beam intensity is just at threshold to maintain the wave stationary. Once formed an absorption wave absorbs all or a large fraction of the incident radiation, thereby shielding the target. It is therefore important to determine the conditions under which absorption waves are ignited. In this paper we will be interested in LSC waves, which occur at rather low incident intensities, and we will be interested in their ignition and propagation at sea level atmospheric pressure.

Preceding page blank

II. CHARACTERIZATION OF LASER SUPPORTED COMBUSTION (LSC) WAVES

Most investigations of LSC (and other absorption) waves have been carried out with laboratory lasers. However, we are interested in the properties of LSC waves under field conditions where the target areas irradiated are much larger than those in the laboratory. Since the incident intensities are comparable, some information obtained in the laboratory is applicable to the field situation. Because opportunities for performing field experiments have been quite limited, and control over beam quality not completely satisfactory we make use of laboratory results to the extent practicable. We have used two high power lasers (HPL No. 1 and HPL No. 2) to produce large scale LSC waves and have used the Ballistic Research Laboratories 2 KW laser and the United Aircraft Research Laboratories (UARL) 6 KW laser in our laboratory investigations. All of these lasers operate CW at a wavelength of $10.6 \mu\text{m}$.

We can best characterize LSC waves by first describing the experimental observations.

HPL NO. 1 EXPERIMENTAL RESULTS. Following are the principal observations concerning LSC waves produced with this laser:

1. There is a bright luminous region ~ 100 cm long and 10 to 15 cm in diameter. Some internal structure is discernible on the film records. In particular, there is a much smaller central region which is much brighter than the remainder of the luminous region. This is the region of intersection of the laser beam with the luminous volume. The length of the brightest region, the LSC wave proper, is a few centimeters. Spectroradiometer results also attest to the fact that the strongly radiating region is a few centimeters long; in addition, we have noted that the shadows formed of nearby objects are sharp indicating that the main portion of the LSC wave radiant emission emanates from a small region.

2. Each LSC wave has its highest velocity when first distinguishable on the framing camera records. (With the instrumentation available for these experiments, the first involving large scale LSC wave behavior, LSC waves could not be distinguished from the surrounding luminous region until they had propagated a considerable distance from the target; that the LSC waves formed much nearer to the surface was clearly proven, however, by spectroradiometer data.) The propagation velocity decays approximately linearly with distance. The LSC wave peters out, i.e., ceases to be luminous, when the propagation velocity has decayed to ~ 1 m/sec.

3. The luminous region behind the wave front has a small upward acceleration.

4. After an LSC wave forms, action--greatly reduced--continues at the same location on the target.

5. The luminous region has sharply defined boundaries. Its structure does not change rapidly; substantial changes occur over periods of tens of milliseconds.

6. Burning particles are invariably present.

7. No LSC waves formed when an airflow over the target was provided; however, the lowest airspeed used was Mach 0.1.

In these experiments intensities were in the range of 20 - 40 KW/cm² for the most part. LSC waves were formed with aluminum and molybdenum targets at intensities of 20 KW/cm²; they were not formed with titanium at this intensity, nor in boron or rubber at 25 - 30 KW/cm², nor in tungsten at 40 KW/cm². An LSC wave was formed in tungsten (during a power excursion of the laser) at an intensity of 90 KW/cm². However, the intensities quoted are only accurate to $\pm 50\%$ owing to inaccuracies both in beam power and irradiated area. Also, the number of irradiations was small, so that definitive conclusions regarding ignition threshold cannot be obtained from these results.

HPL NO. 2 EXPERIMENTAL RESULTS. Since the early experiments we have enjoyed an additional limited opportunity to produce large scale LSC waves. We designed new apparatus for studying LSC wave behavior for the latter experiments. Figure 1 shows schematically a system used to obtain time, space, and wavelength resolved spectra of LSC wave emission on each frame of a framing camera record. In addition, we arranged narrow band pass filters that permitted simultaneous transmission of LSC wave emitted light in two narrow wavelength regions to a framing camera film. This enabled us to distinguish among the several sources of light emission in the LSC wave. It enabled us to separate target vapor emission from air emission and to learn something of the transition from a target vapor dominated LSC wave to an air LSC wave.

Additional observations concerning LSC wave characteristics obtained with the aid of the equipment just described are as follows:

1. The LSC wave forms at or very near the target in a mixture of target vapor, particles, and air. No air spectra are seen

initially. Some spectra seen in the case of a 2024 aluminum target, for example, are: several Al and Mg lines, NaD lines, and AlO bands.

2. Air spectra are first seen at distances varying between 1 cm from the target surface and 10 to 12 cm depending upon the laser beam intensity.

3. The H_{α} line at 6563 Å, owing to water vapor in the air, is an indicator of the transition from a target vapor to an air LSC wave. This transition takes place over a small spatial region during a time of several milliseconds; it takes place on the lower edge of the LSC wave (the laser beam being horizontal).

4. Air spectra always include O, N, and H lines; N^+ lines are seen sometimes, apparently only at the highest intensities reached in the experiments, and only near the wave front.

5. The LSC waves upon occasion split into two or three distinct pieces which sometimes separate by more than one meter; this behavior is associated with the spatial structure of the beam.

6. The threshold for ignition of an LSC wave in aluminum, the material for which we have the largest body of results, is approximately the same as found in the earlier experiments, 20 KW/cm².

Some variation from the above general behavior is contained in the experimental results, probably caused by variations in details of the laser beam behavior. Table 2 summarizes typical characteristics of the spectral results. Figure 2 depicts the behavior in a specific case.

III. PHYSICAL PHENOMENA INVOLVED IN LSC WAVE BEHAVIOR

Guided by the experimental observations, we now wish to provide a description of the physical phenomena involved in LSC wave behavior and, in the following section, a more tentative description of the very important, more complicated initiation mechanisms, concerning which few experimental observations exist.

LSC WAVE TEMPERATURE. After formation an LSC wave is in pressure equilibrium with the surrounding air. Thus, except for minor perturbations, the pressure may be taken as one atmosphere. This, along with the temperature to be calculated, determines the degree of ionization; the latter coupled with the temperature and pressure determines the absorption coefficient for the impinging laser beam. Since, according to experiment, the laser beam is absorbed over a

distance of a few centimeters we can determine a range within which the temperature must lie. Specifically, the relationship connecting the degree of ionization, α , absorption coefficient, K_v , pressure, p , and temperature, T , is

$$\alpha^2 = \frac{(T/10^4)^{7/2} K_v}{10.4 p^2 g}, \quad (1)$$

g being approximately constant ($= 2.3$) over the temperature range of interest (4). This equation, with p and g fixed as stated, and α a tabulated function for air (7,8), is readily solved for assumed values of K_v , which must be in the range of 0.1 to 0.7 to be in accord with the observed absorption in the experiments. Solving Eq. (1) under these conditions, we obtain temperatures in the range 11,500 to 15,200 °K.

For the case of air seeded with target vapor we cannot use the tabulated properties of air. Instead we use the Saha equation for (single) ionization equilibrium. However, since the degree of target vapor ionization is large we have modified the usual form of this equation (6). The calculation of LSC wave temperature proceeds in the same fashion for this case except that we must now assume values for the molar concentration of target vapor. So doing, we find that the temperature, in the case of an aluminum target, lies between 6000 and 8000 °K. (Aluminum vapor provides almost all of the ionization, air a small amount, because aluminum has a much lower ionization potential than any of the constituents of air.)

The calculated temperature ranges are consistent with several experimental observations. First, for a given temperature and LSC wave volume, we obtain the power radiated from calculated curves which have been checked against experiment in the range of interest (9). We find that most of the power absorbed from the laser beam is reradiated. Next, we have calculated spectra emitted by air at temperatures of interest from results of Biberman et al. (10), modified to account for geometrical differences. We find that:

1. The radiating volume is "thin"; i.e., almost all of the emitted radiation escapes the hot plasma except for that emitted at wavelengths shorter than approximately 1100 Å.

2. The spectral distribution of the radiation is consistent with spectroradiometer results (measured power radiated in selected wavelength intervals) for an air LSC wave in the 12,000 to 15,000 °K range and for a target vapor seeded LSC wave ~ 7000 °K.

3. An estimate of the total power radiated based on the calculated spectral distribution is consistent with that calculated from the spectroradiometer data, assuming the spectral distribution to be correct (which is consistent with the results already mentioned).

4. For some of the LSC waves observed in the experiments N^+ spectral lines were seen at times near the wavefront (oscillating on and off, owing to laser power fluctuations); for other LSC waves no N^+ spectra were seen. This, taken in conjunction with known air species concentrations as functions of temperature, suggest that the lower and higher LSC wave temperatures are $\sim 12,000$ and $14,000$ $^{\circ}\text{K}$, respectively (9,11).

5. The temperatures deduced are in the general range observed in more carefully controlled laboratory experiments in which the intensities were comparable while the spot sizes were much smaller and the observed radiation source was the limiting case of an LSC wave, a plasmatron. Using the BRL 2 KW laser (in argon) a temperature of $13,000$ $^{\circ}\text{K}$ has been obtained. Using the UARL 6 KW laser (in air) a temperature of $17,000$ $^{\circ}\text{K}$ has been obtained.

POWER BALANCE. Having determined the relevant temperature range, we can make a power balance in the LSC wave. We have shown that the power radiated constitutes the major portion of that absorbed from the laser beam. For a given temperature we can calculate the stored energy, which turns out to be quite small. (There is also stored energy in the luminous region surrounding the LSC wave, the luminosity being caused by metastable species. We have estimated the density of this region from its rate of rise in the earth's gravitational field; this, and the fact that it is at atmospheric pressure, determine its average temperature, ~ 700 $^{\circ}\text{K}$, and its energy content, which turns out to be quite small.) Since the structure of the wave has been observed to change rather slowly, changes in stored energy have a negligible effect on the power balance. Likewise, calculation of the thermal conductivity of air at the relevant temperatures, and consideration of the dimensions of the LSC waves, show that heat conduction plays only a minor role (6).

Finally, there is a transfer of energy by convection; we cannot estimate this accurately. We obtain a rough estimate by calculating the rate of air engulfment by the LSC wave using experimentally observed dimensions (cross-sectional areas) and propagation velocities, and multiplying this result by the energy gain per unit volume of cold air entering the wave. We find that convective power transfer is not small; however, it is not as large as the power radiated.

PROPAGATION VELOCITY. The final characteristic of an LSC wave (after initiation) which we wish to discuss is its propagation velocity. This, we have observed, decays essentially linearly with distance. Raizer's theory (4) provides an expression for the velocity dependent upon the beam intensity (which varies as the wave propagates toward the laser since the beam is focused). However, this theory is applicable only to small scale LSC waves where heat conduction rather than radiation dominates the energy transfer processes. Application of Raizer's theory to the present LSC wave data results in velocities much smaller than those observed. We wish to emphasize, however, that the propagation velocity is only partially determined by coupled radiation and hydrodynamics: a very essential ingredient is the initial mass and velocity of the hot gas. This is associated with the initiation mechanisms to be discussed in the next section. Briefly, the vapor (and particles) which blow off the target and first form an absorbing region, receive a velocity directed from the target toward the laser in the blowoff process. The mass removed from the target and the time to form an absorbing region depend in some fashion upon beam intensity and target material; (for large beams the area irradiated is probably not important.) In addition, as the absorbing region forms and heats, there is an exodus of gas from the volume (to maintain pressure equilibrium with the surroundings); the interaction of the ejected gas with the target results in additional momentum of the material in the absorbing region, directed away from the target. This effect can be quite large according to (one dimensional) calculations by Hall et al. (11). The dimensions of the initial absorbing region, for a given target material, can vary considerably with incident intensity: according to our most recent experimental observations of LSC waves, in which we placed two narrow band pass filters in parallel in front of a lens and framing camera, the transition from a target vapor to an air LSC wave may take place within 1 or 2 cms of the target or may occur 10 or 12 cms away from the target (for an aluminum target).

The observed, approximately linear, velocity decay is then partly caused by the sharing of momentum with engulfed air and partly caused by diminution in beam intensity as the LSC wave progresses toward the laser. (The latter need not always be true: by focusing the beam in front of the target, the LSC wave can be made to propagate into a region of increasing beam intensity; we have done this but, owing to inadequate control over beam quality, no definitive results have been obtained as yet.) The observed velocity decay rates of the LSC waves support the above statements as we have shown by intercomparing decay rates for LSC waves in air formed with different target materials (6). Except for propagation velocity all of the characteristics of an LSC wave, after the transition from

target vapor to air has been made, are determined solely by the properties of air, the target vapor being too dilute to play a role once the air has been ionized significantly.

IV. MECHANISMS OF INITIATION

Initiation of an LSC wave takes place in some mixture of target vapor, burning target particles, and air. The transition to an LSC wave in air usually occurs within 1 or 2 cms of the target surface; however, when the laser beam intensity is barely sufficient to support an LSC wave in air, the transition may occur up to 10 or 12 cms from the target. In some cases the transition fails to occur, the LSC wave petering out as the target vapor becomes more dilute.

A target vapor-air mixture at the boiling point of most metals does not have a sufficient degree of ionization to absorb the laser beam significantly, thereby become hotter. Consider, for example, the case of aluminum whose boiling point is 2720°K . Pure aluminum vapor at atmospheric pressure does not significantly absorb a $10.6\text{ }\mu\text{m}$ wavelength laser beam below a temperature of approximately 5000°K (12); the mixtures of aluminum vapor and air we previously considered absorbed the laser beam, over the observed distances, at temperatures in the $6000 - 8000^{\circ}\text{K}$ range. We are therefore led to a search for processes which can lead to heating of the electrons (after which these electrons can cause impact ionization resulting in a rate of gain of electrons which exceeds the loss rate; i.e., a cascade process occurs, the medium then becoming capable of absorbing the incident laser beam).

Many mechanisms have been considered for intensities of 10^7 W/cm^2 or higher (13). The levels of interest here are $\sim 10^4$ to 10^5 W/cm^2 . At the higher intensities the transient pressure at the target surface can be much greater than atmospheric with the result that the temperature can greatly exceed the normal boiling point; the thermionically emitted electrons are then characterized by a much higher temperature and give rise to impact ionization. Walters (13) has considered various suggested mechanisms which play a role in this higher intensity regime. All of the suggested mechanisms appear to be applicable only at intensities much higher than those of present concern. On the other hand the role of particles blown off the target, of which a great many are experimentally observed, has not been considered. (We have demonstrated (6) that the small particles keep up with the LSC wave as it forms and moves away from the target.) That the hot burning particles play a role seems clear since they are in the vapor-air mixture where initiation occurs while, as we have seen, initiation does not occur at the target surface.

As is the case at the target surface, there appears to be no way the pressure at the surface of a small particle can be raised sufficiently to increase the particle's boiling point significantly (in spite of the large surface tension such particles possess): a very high pressure is required to raise the temperature appreciably while, because of its small volume (hence mass) to area ratio, a small pressure gradient within the particle will cause its disintegration. Such disintegrations are commonly seen in metal flames (14,15) and appear to occur even more readily when the burning takes place in a laser beam (16). Thus the electrons thermionically emitted from the particles do not possess the requisite high temperature. (Their density can be high within a few particle radii; escape to greater distances is prevented by image charges developed on the particles.)

COMBUSTION OF METALS. Most metals have high heats of combustion which lead to high adiabatic flame temperatures, usually ~ 2000 to 4000°K ; the actual flame temperature is lower owing to losses (17). For example, aluminum has an adiabatic flame temperature of approximately 3900°K (18). There has been a great deal of study of the burning of small metal particles as this is of much practical interest from several standpoints. The oxidation mechanisms of various metals differ depending on many thermophysical and mechanical properties of both the metals and their oxides (15). Again taking aluminum as an example, an Al_2O_3 shell forms outside the oxidizing particle; oxidation of the vapor emitted by the enclosed boiling aluminum takes place at the surface of the oxide layer (18). In this case the flame temperature cannot exceed the oxide boiling point of 3800°K since the oxide decomposes upon vaporization, as it does for most metal oxides (17). (A layer of AlO gas can conceivably form outside the molten Al_2O_3 shell, and this highly absorbs $10.6\ \mu\text{m}$ radiation as do some other metal oxides; however, the surrounding vapor has a high thermal diffusivity, and therefore a large diffusion length, even on a msec time scale. A simple calculation shows that the heat capacity of the vapor, which must be heated along with the absorbing AlO gas, is large enough so that the temperature rise is negligible.)

We believe that a clue as to the means by which high electron temperatures come about is to be found in the behavior of metal flames without the presence of a laser beam. It is well known that electron temperatures can greatly exceed the gas temperature and that electron densities can be orders of magnitude greater than their equilibrium density (19). A plausible explanation for this has been given by Von Engel (19): During the oxidation process, products of reaction are formed in excited states. Electrons

ALLEN, *BRAERMAN, & STUMPFEL

undergo collisions with the excited reaction products, part of the excitation energy being converted to kinetic energy of motion of the electrons. This is illustrated by the reaction



where A, e are an atom (or molecule) and electron, respectively, and the asterisk indicates an excited state. Making use of the principal of detailed balance, Von Engel has calculated cross sections for several such processes and obtains fairly large values in cases involving electronic excitation.

While inverse reactions occur, the flow of chemical energy (i.e., the heat of combustion) results in a net transfer of energy to the electrons causing their mean energy or the electron temperature to be higher than that of the neutral gas. While Von Engel's illustrative calculations do not apply directly to metal-oxidant reactions, it is well known that such reactions are among the brightest chemiluminescent reactions (20). The radiation from such reactions emanates from electronically excited atoms and molecules populated directly by chemical reactions.

The effect of the laser beam's impinging on the oxidizing particles must be to increase the temperature of the entire system (it increases the flame temperature by counteracting heat losses): The absorption of infrared radiation by small particles is generally broadband (21) in contrast to absorption in molecular band systems. This absorption is high even for a "good conductor" such as aluminum--at the high temperatures in question all of the metals have high resistivities.

MATERIAL DEPENDENCE. Most of the LSC waves studied have been produced with metallic targets. We noted in the Introduction, however, that we did not ignite LSC waves in boron or rubber at intensities sufficient for ignition in aluminum and molybdenum. Likewise Fowler et al. (12) have ignited LSC waves in many metals with a 6 KW CO₂ CW laser (at intensities of 10⁶ to 10⁷ W/cm²--higher than those of concern here, yet lower than the intensities at which target plasma properties rather than properties of the solid target are dominant); they failed to ignite wood or plexiglas at the highest intensity they were able to achieve, twice that required to ignite metals. These materials have low flame temperatures. Pyroceram, composed of several metallic oxides, ignited LSC waves with the same intensities as required by the metals. The vaporization temperature of pyroceram is ~ 2000 to 3000 °K. The possibility exists that the metallic oxides decompose, then reoxidize at a later stage in the process,

forming reaction products in excited states. This could lead to heating of thermionically emitted electrons followed by ignition as described above. (In all cases the reaction products remaining in the LSC wave after formation are dissociated into their constituent atoms, absorbing energy in the process—this after they have been instrumental in igniting the LSC wave; we note also that at high temperatures thermionic emission from insulators can be high (22) owing to the large decrease in the bandgap.)

CONVERSION TO AIR LSC WAVE. Finally, the conversion of the LSC wave in the target vapor-particle-air mixture to an LSC wave in air probably takes place in a straightforward manner. The energy absorbed by the LSC wave is transferred to the engulfed air as the wave propagates and the target vapor becomes progressively more dilute. The transition occurs on the lower edge of the wave (for a horizontal beam), air flowing into the wave mainly at the bottom as material from the target vapor-air mixture rises.

V. SUMMARY AND CONCLUSIONS

The principal experimental findings and the conclusions based upon a description of the physical processes which at least qualitatively explain all of the experimental results are:

1. LSC waves in air form anywhere from 1 cm to 10 or 12 cm from the target surface depending upon beam intensity and target material. Typical dimensions of the most intense emission region are: 2 to 3 cm diameter; 3 to 4 cm length.
2. The absorbing region is first formed in a mixture of target vapor, particles, and air.
3. Burning particles play a major role in LSC wave ignition at incident intensities of 10^4 to 10^5 W/cm². In the oxidation reactions electronically excited products form and transfer energy to electrons thus causing an electron temperature substantially higher than the gas temperature. Cascade ionization follows, the electrons giving rise to impact ionization at a rate exceeding the electron loss rate. Details are specific to each target material, though some similarity in the behavior of various metals is to be expected.
4. Metals and some metallic compounds ignite LSC waves at least somewhat more readily than do other materials that have been tried. A low ionization potential material favors ignition.

5. An LSC wave receives initial momentum directed away from the target the magnitude of which depends upon incident intensity and target material.

6. The LSC wave when first formed in the target vapor-air mixture has a temperature generally in the range of 6000 - 8000 °K. The target vapor seeds the air and supplies almost all of the ionization.

7. After the transition from an LSC wave in the target vapor-air mixture to an LSC wave in air, the temperature reaches the 12,000 - 14,000 °K range.

8. The air LSC wave radiates a considerable amount of energy in the vacuum ultraviolet wavelength region of the spectrum. This radiation is to a considerable extent absorbed in the surrounding cold air and plays a role in LSC wave propagation.

9. After formation the LSC wave absorbs a large fraction, in some cases almost all, of the incident laser beam, thus cutting the beam off from the target. When one LSC wave peters out--after reaching a low velocity ambient air currents coupled with diminished beam intensity result in insufficient power to sustain it--another forms near the target, provided the beam characteristics have not changed. Note: No LSC waves have been formed in the few experiments in which we provided air flow over the target. However, the lowest air speed we used in the experiments was Mach 0.1.

REFERENCES

1. Yu. P. Raizer, Soviet Physics JETP 21, 5 (1965).
2. Yu. P. Raizer, Zh ETF Pis' ma 7, 2 (1968).
3. F. V. Bunkin, V. I. Konov, A. M. Prokhorov, and V. B. Fedorov, Zh ETF Pis' Red. 2, 11 (1969).
4. Yu. P. Raizer, Soviet Physics JETP 31, 6 (1970).
5. P. E. Nielsen and G. H. Canavan, AFWL Laser Digest, Dec. 1971.
6. F. J. Allen, O. R. Lyman, J. C. Barb, K. Frank, and C. R. Stumpf, BRL Report No. 1643, April 1973. (Secret).

ALLEN, * BRAERMAN, STUMPFEL

7. A. S. Predvoditelev, E. V. Stupochenko, E. V. Samuilov, I. P. Stakhanov, A. S. Pleshanov, and I. B. Rozhdestvenskii, Tables of Thermodynamic Functions of Air for the Temperature Range 6000 - 12,000 °K and Pressure Range 0.001 - 1000 atm., Infosearch Limited, London, 1958.
8. J. Hilsenrath, AEDC TN 56-12, Addendum.
9. J. W. Bond, Jr., K. M. Watson, and J. A. Welch, Jr., Atomic Theory of Gas Dynamics, Addison-Wesley, Reading, Massachusetts, 1965.
10. I. V. Avilova, L. M. Biberman, V. S. Vorobjev, V. M. Zamalin, G. A. Kobzev, A. N. Lagar'kov, A. Ch. Mnatsakanian, and G. E. Norman, J. Quant. Spectrosc. Radiat. Transfer 9, 1969, pp. 89-122 and Tables, pp. 1285-1312. (Articles in Russian).
11. R. B. Hall, W. E. Maher, and P. S. P. Wei, AFWL-TR-73-28, June 1973.
12. M. C. Fowler, D. C. Smith, C. O. Brown, and R. J. Radley, Jr., UARL Report N921716-7, Jan. 1974, ARPA Order No. 2113.
13. C. T. Walters and R. H. Barnes, Battelle Columbus Laboratories, Nov. 1973, ARPA Order No. 2113.
14. R. Friedman and A. Macek, Combustion and Flame 6, 1962, pp. 9-19.
15. T. A. Brzustowski and I. Glassman, Progress in Astronautics and Aeronautics, Vol. 15, 1964, pp. 41-74.
16. I. Liebman, J. Corry, and H. E. Perlee, Combustion Science and Technology 5, 1972, pp. 21-30.
17. A. V. Grosse and J. B. Conway, I and EC, 50, 4, 1958.
18. R. W. Bartlett, J. N. Ong, Jr., W. M. Fassell, Jr., and C. A. Papp, Combustion and Flame 7, 1963, pp. 227-234.
19. A. Von Engel, Brit. J. Appl. Phys. 18, 1967, pp. 1661-1677.
20. S. E. Johnson, P. B. Scott, G. Watson, Xonics, Inc., Final Rept., Contr. No. DAAH01-73-C-0548.
21. H. C. Van de Hulst, Light Scattering by Small Particles, John Wiley and Sons, Inc., New York, 1957.
22. S. L. Soo, J. Appl. Phys. 34, 6 (1963).

Table 2. Spectral Observations

Spot Size	Strongest Observed Spectra	Physical Interpretation	Remarks
3-12 cm	AlO Molecular bands; Strong Al lines; impurity lines (Na and Mg)	Vaporization of aluminum; chemical combustion of Al vapor and particles	Emission often extends many cms from target
2-6	Very strong Al line emission and Na, Mg line emission	Vaporization; excitation of Al (and impurities)	Emission extends a few cms from target; many KW of power radiated in visible region
2-4	N, H, O lines and occasional N+ lines	Ionization of the air	LSC wave propagates away from target

Aluminum target

Laser beam power \approx 130 KW

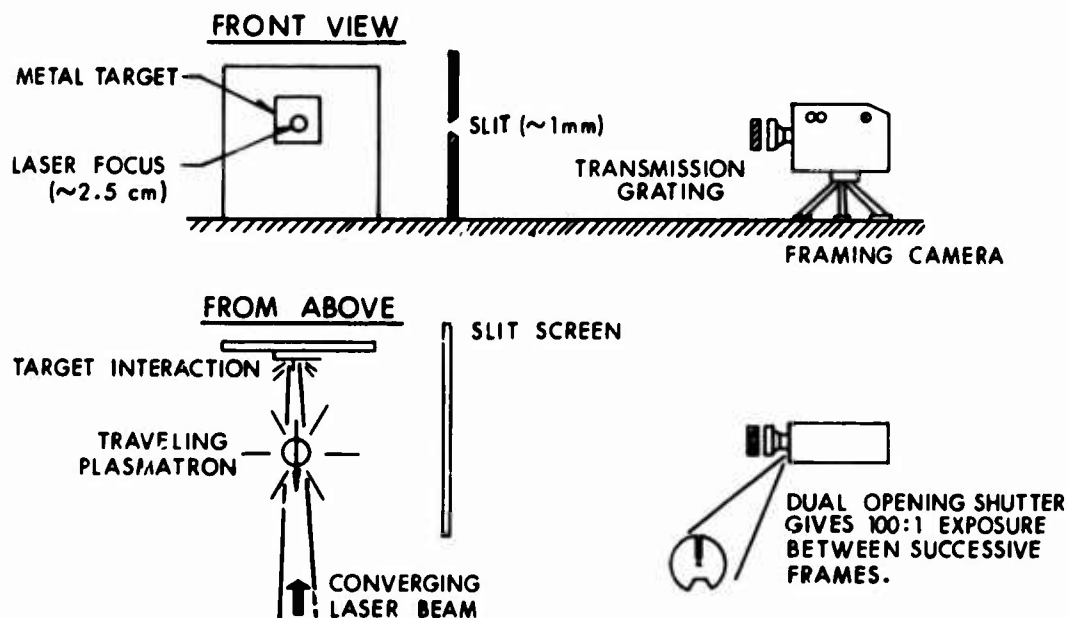


Fig. 1. Transmission Grating Measurement of LSC Wave Emission.

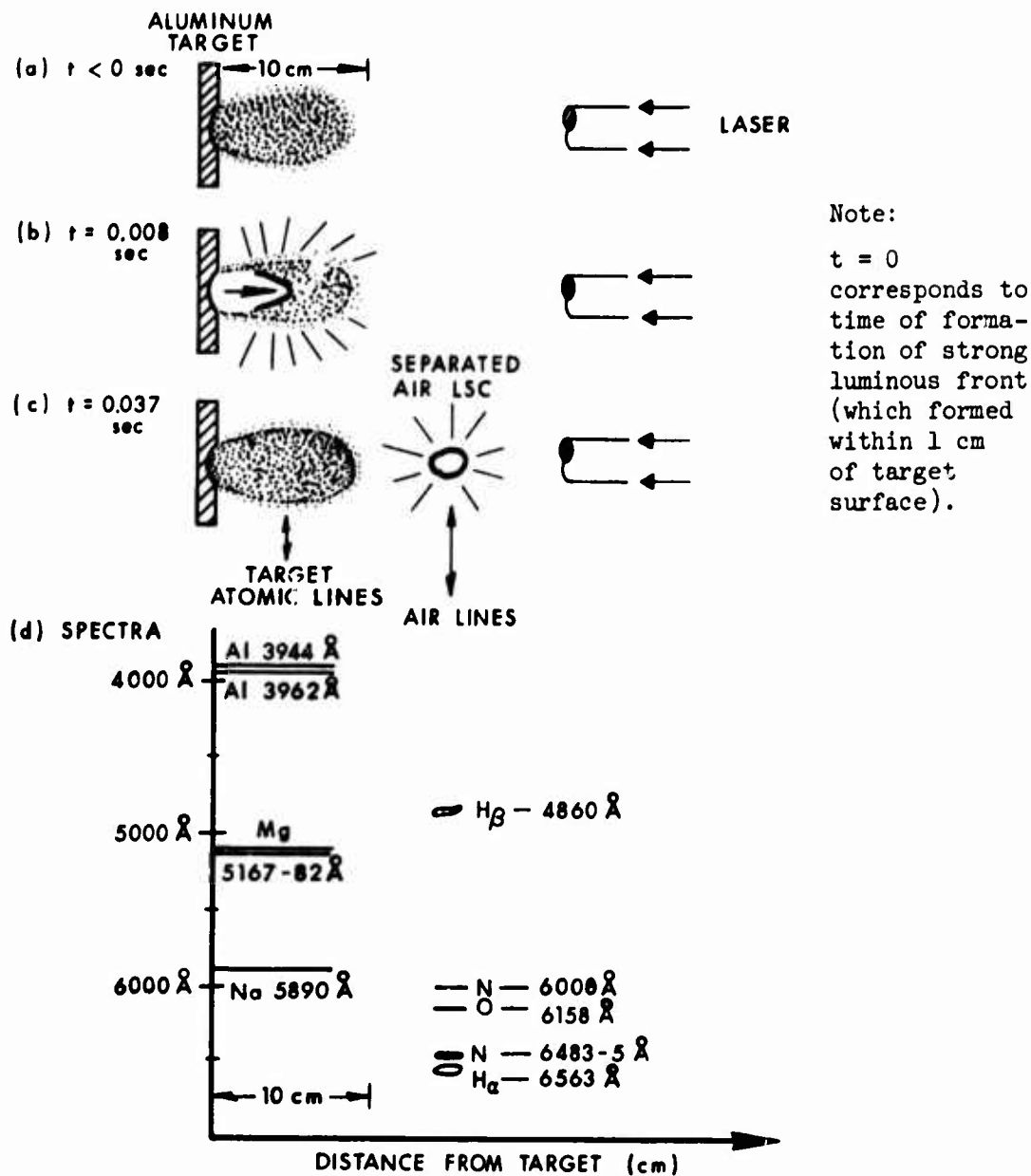


Fig. 2. Sketches of Sequence of Events Leading to LSC Wave Formation and Propagation. (a) Luminous target vapor and oxide particles. (b) Strong luminous front 0.008 sec after formation. (c) Strongly radiating air LSC wave separated from target vapor. (d) Spectra corresponding to (c); strongest emission lines are shown and serve to distinguish between air and target vapor dominated behavior.

**ELECTROSTATIC MEANS FOR INTRUSION
DETECTION AND RANGING**

**ALAN D. ARONOFF, MR.
WILLIAM H. BOGHOSIAN, DR.
HOWARD A. JENKINSON, MR.
FRANKFORD ARSENAL
PHILADELPHIA, PA**

INTRODUCTION

Electrostatic intrusion detection techniques have been utilized in the past for a variety of applications, both military and civilian. The basis for these techniques has been essentially the measurement of the distortion of an electric field due to the presence and/or motion of an object. An equivalent point of view results from analyzing the electric field perturbations in terms of the self and mutual capacitances of the detector probe and the object, and the variations of these capacitances with changes in the relative location of the detector and object.

A passive electrostatic detection system is one in which the electric field is generated external to the detector or sensor. If, however, the sensor itself generates the electric field which is perturbed by the object, the system is known as an active electrostatic detector. In this paper we deal with a passive detection system in which the earth's electric field, which is normally directed vertically to the earth's surface, is perturbed or distorted by the presence of an intruder.

The magnitude of the distortion depends on the relative position of the intruder and sensor and the configuration and dimensions of both. Fig. 1a shows the undisturbed field with the sensor depicted as a probe or antenna of arbitrary shape connected to a meter. It is assumed that probe size and shape are such that it does not disturb the field being measured. Fig. 1b shows the field distorted by the presence of an intruder modelled in this case as a grounded conductor.

Preceding page blank

ARONOFF, BOGHOSIAN and

JENKINSON

The entire intrusion detection system is shown in block diagram form in Fig. 2.

The above model represents an oversimplification of the effect of a walking intruder on the earth's electric field. Such an intruder cannot be represented simply as a moving grounded conductor. A person walking or scuffing may develop potentials of over 1000 volts depending on shoe sole thickness and resistance and on the nature of the ground surface and material. Moreover, as he raises one foot at each step his body capacitance to ground reaches a minimum and this in turn causes his potential to rise to a maximum. When both feet are momentarily making ground contact, the charge potential is lowest, and, thus, there results a characteristic rise and fall in potential due to walking. This in turn introduces a similar disturbance in the earth's field in the vicinity of the moving person. This phenomenon is readily observed experimentally and it can be shown that the charge-discharge phenomena of a walking person will produce significant signals at useful ranges.

This paper is based on the results of an initial study to determine the feasibility of a reliable, reasonably inexpensive electrostatic intrusion detection system for field use. This initial effort has been concerned mainly with carrying out an analytic program through modelling and computer studies together with a preliminary experimental program based on these studies. In subsequent effort it is intended to expand these studies to include signal processing techniques required for the optimization of the detection and identification performance of the system and, in addition, breadboard components suitable for application of this detection technique to scatterable mine systems.

MODEL STUDIES

In this section we will first consider distortions in the earth's electrostatic field due to grounded and charged intruders. We will next introduce a time dependence due to motion toward the sensor as well as the charge/discharge phenomena attributable to the raising and lowering of the intruder's feet. We then outline a system capable of detecting these distortions through atmospheric coupling. The effects of finite input resistance and non-zero input capacitance of the detector are also accounted for.

To determine the distortion, which we define as the difference in the field due to the presence of an intruder after subtracting out the naturally occurring earth's field, it is necessary to solve Laplace's equation in a geometry appropriate to that of the intruder.

ARONOFF, BOGHOSIAN and

JENKINSON

If the intruder is modelled as a conducting surface of constant value in some orthogonal coordinate system in which the Laplacian operator separates into ordinary differential operators, then it is not too difficult to obtain a closed form analytic solution.

We first consider analytic solutions for some simple geometries which approximate the intruder. The simplest of all is the grounded hemisphere (Fig. 3). It is well known that the potential outside and above a grounded hemispherical boss on a infinite grounded plane in a uniform electric field is given by⁽¹⁾

$$V(\rho, z) = E_0 z \left\{ 1 - a^3 (\rho^2 + z^2)^{-3/2} \right\}, \quad a = \text{radius of boss} \quad (1)$$

where E_0 is the constant, undisturbed earth's electric field, 100 v/m in this calculation.

However, the hemisphere is a rather poor model since it implies that the intruder is as wide as he is tall. A much better model is a grounded hemiprolate-spheroidal (half watermelon) boss on the infinite plate (Fig. 4). By switching to the prolate spheroidal coordinate system ζ, η defined by the transformations:

$$\rho^2 (\eta^2 - 1)^{-1} + z^2 / \eta^2 = C_2^2 \quad (2)$$

$$\rho^2 (\zeta^2 - 1)^{-1} + z^2 / \zeta^2 = C_2^2 \quad (3)$$

where C_2 is determined from the semi-major axis, C , and semi-minor axis, b , of the spheroid by $C_2 = \sqrt{C^2 - b^2}$, it can be shown⁽¹⁾ that V is now given by

$$V = E_0 z \left[1 - \frac{\frac{1}{2} \ln((\eta+1)/(\eta-1)) - 1/\eta}{\frac{1}{2} \ln((\eta_0+1)/(\eta_0-1)) - 1/\eta_0} \right], \quad (4)$$

where $\eta_0 = C/C_2$. In these models we used $b = 0.15$ m and $C = 2.0$ m which approximate the dimensions of a man.

There are other models which may represent limiting values for the intruder. For example, we considered the grounded, half-elliptic cylinder of semi-major axis C and semi-minor axis b , shown in Fig. 5. The solution of Laplace's equation for this geometry can be shown to be⁽²⁾:

$$V = E_0 z \left\{ 1 - \frac{\eta_0 (\eta_0 + \sqrt{\eta_0^2 - 1})}{\eta (\eta + \sqrt{\eta^2 - 1})} \right\} \quad (5)$$

where η and η_0 are the same as in the above prolate spheroid (Eqns. 2 and 3).

We can also consider the generalized ellipsoid with no axial symmetry (Fig. 6). Following Jeans⁽³⁾, if one has the ellipsoid

ARONOFF, BOGHOSIAN and

JENKINSON

$$x^2/a^2 + y^2/b^2 + z^2/c^2 = 1 \quad (6)$$

then the conicoid

$$x^2(a^2 + \theta)^{-1} + y^2(b^2 + \theta)^{-1} + z^2(c^2 + \theta)^{-1} = 1 \quad (7)$$

will be confocal to the ellipsoid for all values of θ .

Solving the cubic equation for θ gives the three values λ , μ , ν for each cartesian point x , y , z where $\lambda = \text{const.}$ is a family of ellipsoids, $\mu = \text{const.}$ is a hyperboloid of one sheet, and $\nu = \text{const.}$ is a hyperboloid of two sheets. If we require $a > b > c$, and assume the field is in the x direction, it can be shown that

$$V = E_0 \left[\int_{\lambda_0}^{\lambda} \frac{d\lambda}{(a^2 + \lambda)(b^2 + \lambda)(c^2 + \lambda)} \right] \bigg/ \int_{\lambda_0}^{\infty} \frac{d\lambda}{(a^2 + \lambda)(b^2 + \lambda)(c^2 + \lambda)} \quad (8)$$

where λ_0 is the coordinate of the intruding ellipsoid of height a , length b , and breadth c , and

$$\Delta_{\lambda} = \sqrt{(a^2 + \lambda)(b^2 + \lambda)(c^2 + \lambda)}.$$

Thus, the integrals in Eqn. (8) are of the form

$$I = \int_{\lambda_0}^{\lambda} \frac{d\lambda}{(a^2 + \lambda)^3 (b^2 + \lambda)(c^2 + \lambda)} \quad (9)$$

which can be evaluated in terms of appropriate elliptic integrals of the first and second kinds⁽⁴⁾.

We have plotted the field distortions of the prolate spheroid and elliptic cylinder geometries in Fig. 7. The field distortion was obtained directly from the potential distribution matrix $V(i,j)$, where $V(i,j)$ represents the potential at the point $Z(i)$ and $\rho(j)$ and then using a central difference formula to approximate the gradient operator:

$$E_z(i,j) = (\Delta z)^{-1} \{ V(i+1,j) - V(i-1,j) \} - E_0 \quad (10a)$$

$$E_{\rho}(i,j) = (\Delta \rho)^{-1} \{ V(i,j+1) - V(i,j-1) \} \quad (10b)$$

where $\Delta z = [z(i+1) - z(i-1)]$ and $\Delta \rho = [\rho(j+1) - \rho(j-1)]$.

It is noted that the horizontal components of the field distortion are much smaller than the vertical components, except perhaps at points very close to the intruder. Furthermore, at horizontal distances from the intruder greater than 3 m, E_{hor} is proportional to the height off the ground, while E_{vert} is almost height independent. It would then seem desirable to detect the vertical component, taking advantage of these two important features.

Note further that the distortions for the elliptic cylinder are much stronger and fall off slowly as a function of ρ . Thus, we

ARONOFF, BOGHOSIAN and

JENKINSON

put this as a maximum upper bound for the distortion due to a grounded object of elliptical cross section.

It is an oversimplification, however, to assume that the intruder remains at ground potential. For example, an initially grounded, conducting intruder has a charge distribution induced on his surface by the external field to enable him to remain at ground potential. Should the intruder break the connection to ground by lifting his feet, he reduces his capacity to ground which increases his potential. A similar if weaker effect occurs when the intruder picks each leg up, especially if his resistance to ground is not particularly high. Furthermore, scuffing shoes on the ground can lead to static charge being accumulated on the body, which must leak away in a finite time. The amount of charging is not inconsiderable. Over 20 years ago the Bureau of Mines⁽⁵⁾ investigated the hazard of static electricity in operating theaters where there was danger of igniting explosive anesthetics. In a very comprehensive study it was shown that, depending upon the nature of the clothes, linens, furniture surfaces, shoes, and humidity in the room, potentials of 500-18000 volts could be generated on human beings. This charge leaked away through shoes with resistance of 10^6 - 10^9 ohms depending on the shoe material, which varies from partially conducting leather to highly insulating rubber. The time dependence of the potentials was also considered in that report. Thus, it becomes very necessary to obtain the field distortion due to a charged body. For this calculation we assume that the charged body is sitting on a thin insulating pad just above the ground.

Since the ground plane is at zero potential, one may resort to the theory of images to solve this problem. The application of this method to a charged conducting hemisphere on a ground plane leads to the well known exterior solution

$$V(r, \theta) = V_0 \left[3/2 \left(\frac{a}{r} \right)^2 P_1(\cos \theta) - 7/8 \left(\frac{a}{r} \right)^4 P_3(\cos \theta) + (11/16) \left(\frac{a}{r} \right)^6 P_5(\cos \theta) - \dots \right] \quad (11)$$

where $P_1(\cos \theta)$ is the Legendre polynomial of the first kind. Again while the sphere is not a good approximation to the man, we can still use the same technique to solve the problem of the charged conducting prolate ellipsoid.

The general solution of the Laplacian in prolate spheroidal coordinates with azimuthal symmetry is⁽³⁾

$$V(\zeta, \eta) = \sum_i \left\{ A P_i(\zeta) + B Q_i(\zeta) \right\} \left\{ A' P_i(\eta) + B' Q_i(\eta) \right\} \quad (12)$$

where Q_i is the i th Legendre polynomial of the second kind. We impose the boundary requirements:

ARONOFF, BOGHOSIAN and

JENKINSON

but finite (10^{14} ohm) resistance. Using these values and the experimentally measured response time, the probe resistance was calculated to be about 10^{11} ohms. The same experiment was simulated on the computer using the above values of the model parameters in the limit of the probe capacitance approaching zero. The results were quite comparable.

Sensor response to the input signal shown in Fig. 10 has been calculated for various values of the three sensor parameters. The output for two combinations of these is shown in Fig. 12. The smoothing effect of the larger electrometer input capacitance is evident. It is possible that under favorable ionization conditions, the probe resistance would fall to 2×10^{10} ohms. A model calculation using this value was made. The results, plotted in Fig. 13, clearly show the enhancement of structure in the sensor output signal.

EXPERIMENTAL STUDIES

To evaluate the feasibility of employing electrostatic means as an intrusion detection method, various state-of-the-art electrometers were considered with respect to their input resistance, capacitance and input offset current. When run open circuit, offset current will charge the probe and quickly saturate the electrometer. Due to its high input impedance and its particularly low offset current (less than 3×10^{-15} amp), the Princeton Applied Research Corp. PAR-135 electrometer was selected for further experimental work.

Two probe types were tested. One was simply a straight length of stainless steel tubing, 0.037 inch in diameter, projecting 10 cm from the input jack of the PAR-135. The other type was made from the same material projected 5 cm from the PAR-135, and had a small amount of Polonium fastened to the end.

Polonium ionizes the air in its immediate vicinity. The ionization provides a relatively rapid means for the probe to come to the same potential as the air around it, effectively decreasing the probe resistance. Decreased response time and increased sensitivity are the principal benefits obtained by its use. The enhancement produced by the Polonium is evident in the field test results discussed below.

Several field tests were made to determine the best sensor configuration to utilize in recording the response of these sensors to the intrusion of a person. To minimize possible electrostatic interference, these tests were conducted in open country fields.

To obtain an estimate of local field variations, two sensors

ARONOFF, BOGHOSIAN and

JENKINSON

$$V(\zeta, \eta \rightarrow \infty) = 0; \quad V(\zeta, \eta_0) = \begin{cases} V_0, & 0 \leq \zeta \leq 1 \\ -V_0, & 0 \leq \zeta \leq -1 \end{cases}$$

Outside the spheroid only $P_1(\zeta)Q_1(\eta)$ is finite as $\eta \rightarrow \infty$, so Eqn. 12 reduces to

$$V(\zeta, \eta) = \sum_i A_i P_i(\zeta) Q_i(\eta) \quad (13)$$

where the Fourier coefficients are determined by

$$A_i = (2i+1) V_0 [Q_i(\eta_0)]^{-1} \int_0^1 P_i(\zeta) d\zeta \quad (14)$$

Thus,

$$V(\zeta, \eta) = V_0 \left\{ 3/2 P_1(\zeta) \frac{Q_1(\eta)}{Q_1(\eta_0)} - 7/8 P_3(\zeta) \frac{Q_3(\eta)}{Q_3(\eta_0)} + 11/16 P_5(\zeta) \frac{Q_5(\eta)}{Q_5(\eta_0)} \dots \right\} \quad (15)$$

If the charged body is in a uniform externally applied field we simply superimpose the solutions of Eqn. 4 and Eqn. 15. We have numerically evaluated the potential distribution for several values of V_0 both in and out of the earth's field. The vertical distortions are shown in Fig. 8. It may be seen that if the charge on the body is appreciable, then the distortion due to the earth's field is rather negligible compared to the distortion due to the charged body itself.

Time dependence may be included in the calculation in a straightforward way by considering the two effects described above. First, as the intruder approaches the sensor, the field distortion at the sensor will increase due to his increasing proximity. Second, as the intruder lifts his foot to walk, his potential will rise causing a change in field distortion at the sensor. The combination of these two effects produces a fluctuating field distortion at the sensor, which is considered to be the sensor input signal.

A model for the potential profile of the intruder was obtained in the following manner. Several people, each in turn holding a wire connected to an electrometer, walked slowly across the laboratory. Each person's potential profile was recorded. One of these profiles, shown in Fig. 9, was selected as the basis for the walking model. Note that the subject carries a negative potential due to his retention of electrons as he lifts his feet. Also, the large potential rise is due to the capacitance change as the subject lifts his foot. Smaller changes are attributed to the heel and toe motion as the subject shifts his weight.

For ease in calculation we simplified the profile of Fig. 9 by taking the profile to be periodic once walking commenced, and that local maxima and minima could be connected by straight lines. Vertical field distortion curves were plotted for the six intruder potentials corresponding to the local extrema. Using a measured velocity of .83

ARONOFF, BOGHOSIAN and

JENKINSON

meter/sec, a period of 1.6 seconds, and an initial displacement of ten meters from the sensor, the person's displacement from the sensor at each potential maximum or minimum was calculated and marked on the appropriate distortion curve. Connecting these points with straight lines yielded the idealized triangular wave taken as the walking intruder sensor input model (Fig. 10).

The coupling of the field distortions through the atmosphere to the sensor is a complex process. An isolated electrically neutral wire placed in a potential gradient will distort the field such that it averages the field along its length. If the wire be connected to an electrometer, however, the small input capacitance to ground, leads to additional charge redistribution. Furthermore, the electrometer input resistance is not infinite so some leakage to ground will occur. The result of these phenomena is that the potential on the wire is lowered, or the effective height of the antenna is reduced.

In addition, there is a general atmospheric relaxation time for electrical phenomena due to the low but non-vanishing atmospheric conductivity. In a discussion given by Chalmers⁽⁶⁾, it is shown that if a conductor of area A carries a charge Q and is exposed to the atmosphere, it leaks charge as given by Ohm's law $i = \lambda EA$. Close to the surface $E = Q/A\epsilon_0$; thus, $Q = \lambda Q/\epsilon_0$ and $Q = Q_0 e^{-t/\tau}$ where $\tau = \epsilon_0/\lambda$ is the atmospheric relaxation time and λ is the atmospheric conductivity.

This relaxation time suggests as a model for the atmospheric coupling a parallel resistance capacitance combination which is connected to the electrometer. Thus, a reasonable candidate for an equivalent circuit is given in Fig. 11. For the output of the circuit to reflect the input with little lag, the time constant of the probe resistance and electrometer capacitance, τC , should be reasonably small. In addition, to reduce the voltage drop across the probe resistance, it should be small with respect to the detector resistance, R . Since R is specified to be 10^{14} ohms for state-of-the-art solid state (MOS/FET) electrometers, these conditions can be met by use of a radioactive probe which will reduce the probe resistance to 10^{10} - 10^{12} ohm.

To obtain a value for the resistance of an ionization enhanced probe, an experimental sensor consisting of a state-of-the-art lab type electrometer with an enhanced wire probe was placed in a Faraday cage, and its response to a unit step change in electric field strength was recorded. The resulting curve was similar to that associated with the voltage rise across a capacitor placed in a series with a resistor and the input voltage. The electrometer is specified to have a capacitance of under five picofarads shunted by a very high

ARONOFF, BOGHOSIAN and

JENKINSON

but finite (10^{14} ohm) resistance. Using these values and the experimentally measured response time, the probe resistance was calculated to be about 10^{11} ohms. The same experiment was simulated on the computer using the above values of the model parameters in the limit of the probe capacitance approaching zero. The results were quite comparable.

Sensor response to the input signal shown in Fig. 10 has been calculated for various values of the three sensor parameters. The output for two combinations of these is shown in Fig. 12. The smoothing effect of the larger electrometer input capacitance is evident. It is possible that under favorable ionization conditions, the probe resistance would fall to 2×10^{10} ohms. A model calculation using this value was made. The results, plotted in Fig. 13, clearly show the enhancement of structure in the sensor output signal.

EXPERIMENTAL STUDIES

To evaluate the feasibility of employing electrostatic means as an intrusion detection method, various state-of-the-art electrometers were considered with respect to their input resistance, capacitance and input offset current. When run open circuit, offset current will charge the probe and quickly saturate the electrometer. Due to its high input impedance and its particularly low offset current (less than 3×10^{-16} amp), the Princeton Applied Research Corp. PAR-135 electrometer was selected for further experimental work.

Two probe types were tested. One was simply a straight length of stainless steel tubing, 0.037 inch in diameter, projecting 10 cm from the input jack of the PAR-135. The other type was made from the same material projected 5 cm from the PAR-135, and had a small amount of Polonium fastened to the end.

Polonium ionizes the air in its immediate vicinity. The ionization provides a relatively rapid means for the probe to come to the same potential as the air around it, effectively decreasing the probe resistance. Decreased response time and increased sensitivity are the principal benefits obtained by its use. The enhancement produced by the Polonium is evident in the field test results discussed below.

Several field tests were made to determine the best sensor configuration to utilize in recording the response of these sensors to the intrusion of a person. To minimize possible electrostatic interference, these tests were conducted in open country fields.

To obtain an estimate of local field variations, two sensors

ARONOFF, BOGHOSIAN and

JENKINSON

were placed on the ground. Sensor A was equipped with the 5 cm Polonium enhanced probe; Sensor B, 18 m away, with the 10 cm wire probe. Variations in the earth's electric field were monitored for twenty minutes by both sensors. Such disturbances as did occur appeared on both outputs, but more strongly on the output of the sensor with the Polonium probe, and were probably due to small clouds of ionized gas drifting over the field.

Sensor response to an intruder walking slowly toward it is shown in Fig. 14 for the enhanced probe and in Fig. 15 for the plain probe. The intruder's steps can be detected at about 8 meters with the enhanced probe, and at about $6\frac{1}{2}$ meters with the plain probe. The output signal is stronger also for the enhanced probe, having changed by 170 mv at 3 m. while only by 100 mv at 3 m. for the plain probe.

In these measurements a bias was placed on the recorder in order to suppress the constant earth's field. Therefore, the zero volt reference line is not identified on these recordings, but it may be taken to be the line generated before the intruder nears the sensor, since only intruder induced distortions are of interest for comparison with the model calculation.

Plotting the computed sensor outputs on a linear scale, we find good agreement with the measured response. One of these linear plots, corresponding to an electrometer input capacitance of 1 pf is shown in Fig. 16. Although there is a scale factor of two between the voltage scales of Figs. 16 and 14, there is good correlation in the signature and rate of signal increase as the intruder approaches the sensor. There are several reasons for the presence of the scale factor, the most important of which is the fact that the potential profiles generated in the laboratory did not match the one used in the field.

Sensor response to the static presence of an intruder was approximated by instructing the intruder to pause five seconds between each step taken toward the sensor (See Fig. 17). Each step consists of a small signal as the intruder flexes his legs in preparation for the step, a large spike as he takes the step, and a second small signal as he relaxes. The signal is then fairly constant until the next step, but offset from its previous value. The offset is due only to the proximity of the intruder to the detector.

The last figure, Fig. 18, shows sensor response when the intruder was running in place at various distances from the sensor. It was determined the person used as the intruder in these experiments had very low resistance to ground. Any static charge built up by

ARONOFF, BOGHOSIAN and

JENKINSON

walking would quickly leak off, leaving his body potential close to zero. By running in place, he could increase his resistance to ground and build up his potential, which would increase the effect on the sensor. Detection is easier to recognize in this case also because of the periodicity of the signal. As shown in the figure, he was detected by this sensor at 16 meters.

CONCLUSIONS

The results of preliminary experimental tests appear to verify analytical studies indicating that the characteristic signature of walking persons can be detected by sensors utilizing standard portable solid state electrometers equipped with simple 5 to 10 cm wire antennas. Field test data show detectability to ranges of 8 meters and more.

Future work will be directed toward further studies to establish the characteristics of background noise, optimum antenna configurations, and weather degradation. In addition, signal processing techniques will be studied to optimize sensor performance.

REFERENCES

1. W.R. Smythe, "Static and Dynamic Electricity", 2nd Edition, (McGraw-Hill, 1950).
2. P.M. Morse and H. Feshbach, "Methods of Theoretical Physics", (McGraw-Hill, 1953).
3. James Jeans, "Electricity and Magnetism", Fifth Edition, (Cambridge Press University Press, 1951).
4. I.S. Gradshteyn and I.W. Ryzhik, "Table of Integrals, Series and Products", (Academic Press, 1965).
5. P.G. Guest, V.W. Sikora, and B. Lewis, "Static Electricity in Hospital Operating Suites", Bureau of Mines Report of Investigations #4833, January 1952.
6. J.A. Chalmers, "Atmospheric Electricity", (Pergamon Press, 1957).

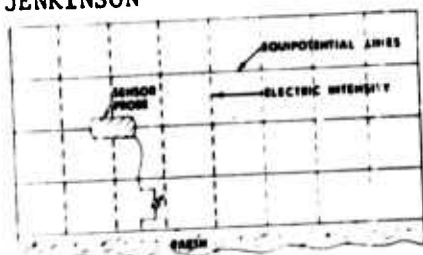


Fig. 1a. Undisturbed Earth's
Electric Field

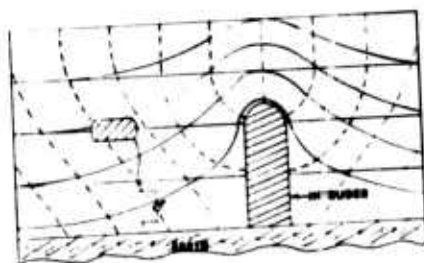


Fig. 1b. Disturbance of Earth's
Field Due to Grounded
Intruder

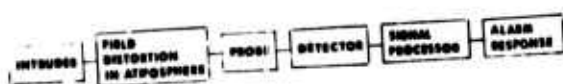


Fig. 2. Block Diagram of Intrusion
Detection System

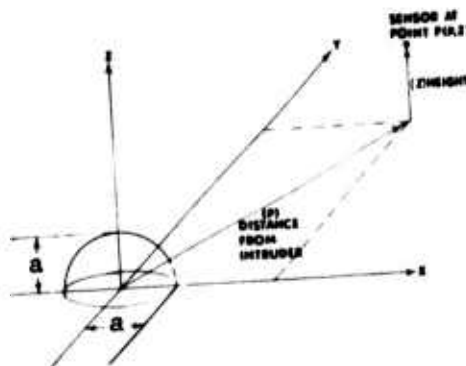


Fig. 3. Hemispherical Intruder Model

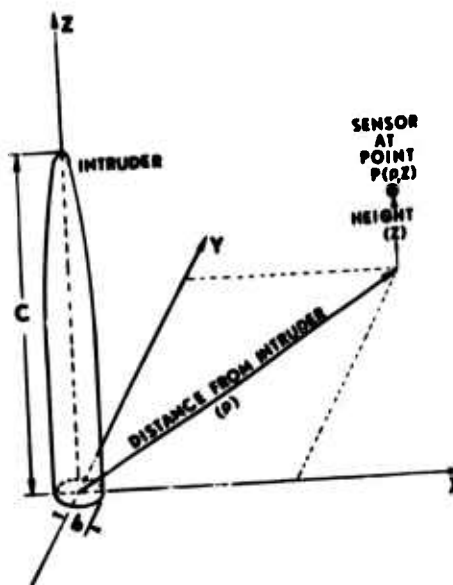


Fig. 4. Hemi-prolate Spheroid
Intruder Model which is
Closer to the True Pro-
portions of a Person

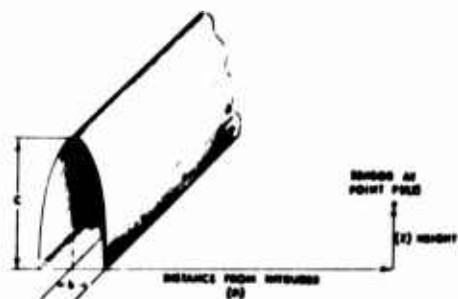


Fig. 5. Hemi-Elliptic Cylinder
Representing an Upper
Limit to Intruder Size

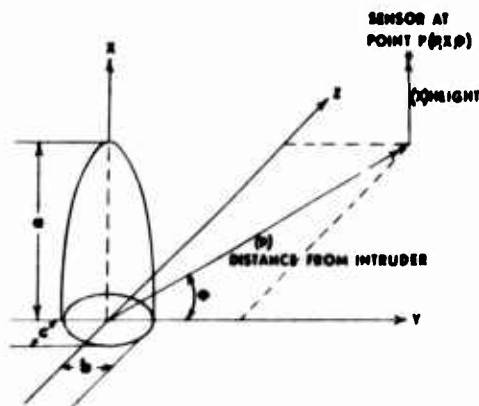


Fig. 6. Hemi-Ellipsoid which
allows for the Intruder's
Lack of Rotational
Symmetry

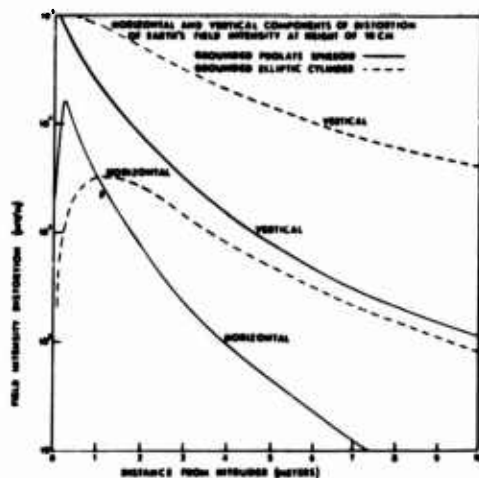


Fig. 7. Comparison of Field Dis-
tortions Due to Grounded
Prolate Spheroid and
Elliptic Cylinder

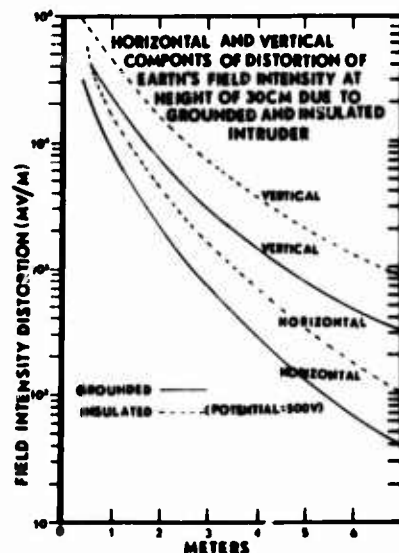


Fig. 8. Comparison of Field Dis-
tortions Due to Grounded
and Insulated Intruder
Using the Prolate Spheroid
Model

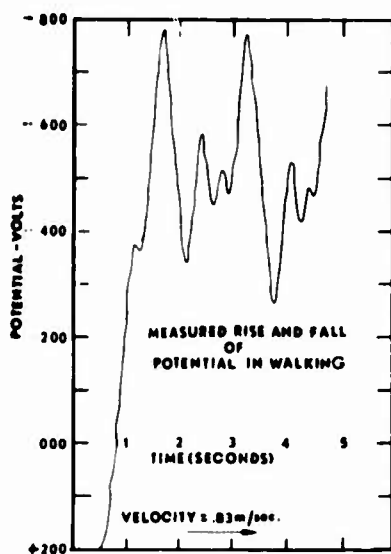


Fig. 9. Laboratory Measurement of the Potential Profile Due to the Rise and Fall of the Feet While Walking

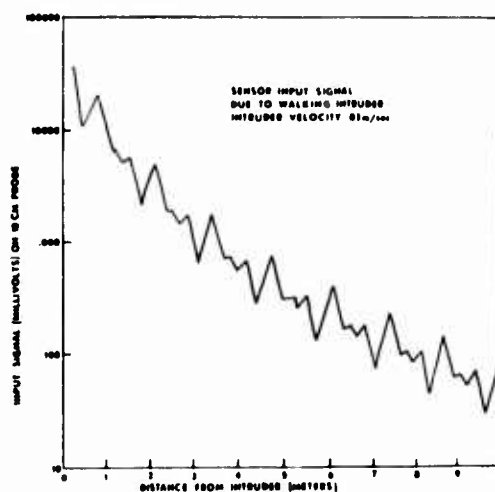
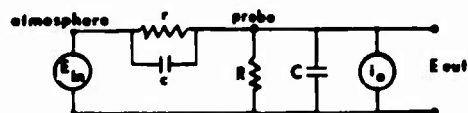


Fig. 10. Input Signal Due to Walking Intruder



E_{in} : sensor input signal
 r : probe resistance
 R : electrometer resistance
 C : electrometer capacitance
 c : probe capacitance
 E_{out} : electrometer output signal
 i_o : input offset current

Fig. 11. Equivalent Circuit for Sensor Input Signal, Atmospheric Coupling, and Detector Input Impedance

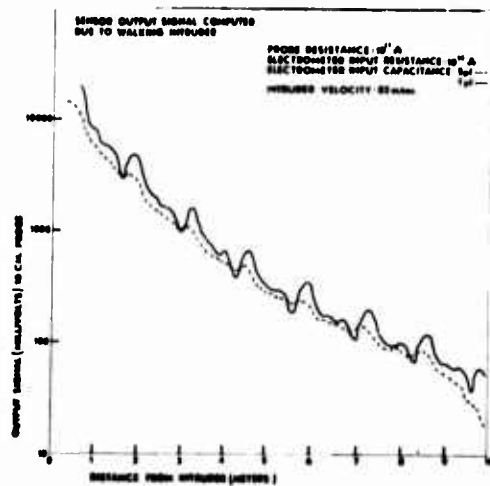


Fig. 12. Computed Sensor Output
for Two Values of Elec-
trometer Capacitance
Using an Enhanced Zero
Capacitance Probe

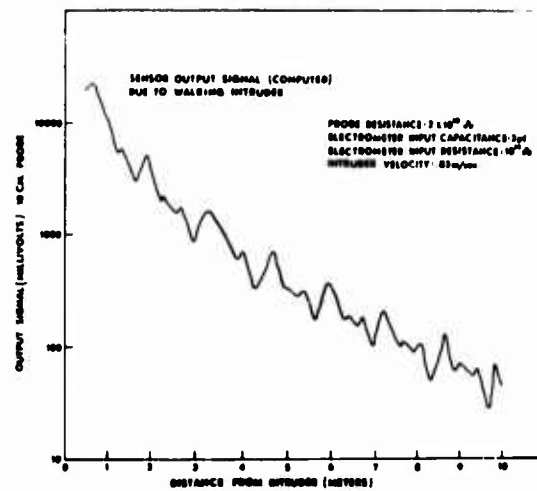


Fig. 13. Computed Sensor Output
Signal for a Highly
Enhanced Probe

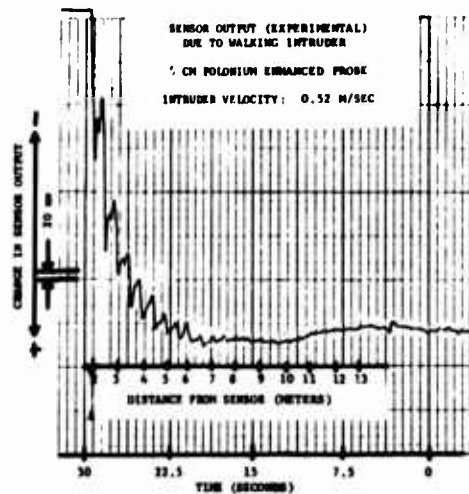


Fig. 14

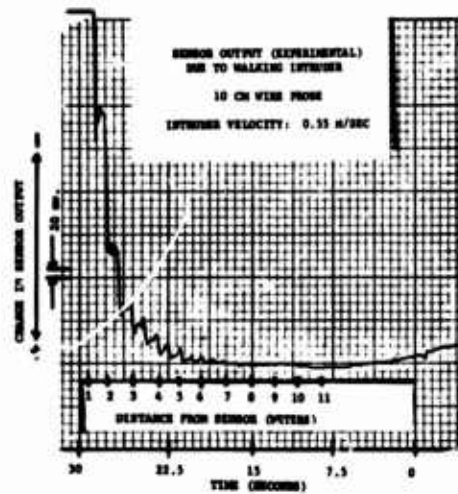


Fig. 15

ARONOFF, BOGHOSIAN and
JENKINSON

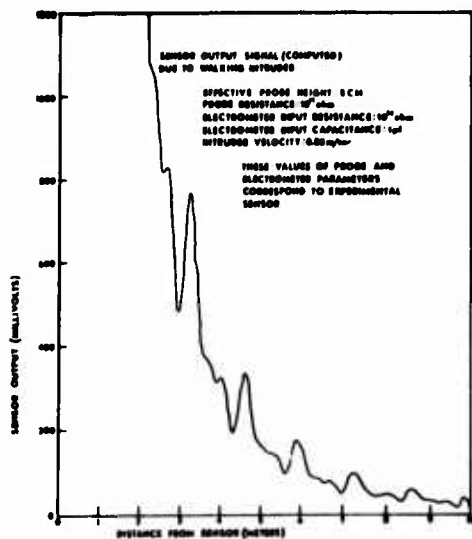


Fig. 16. Computed Sensor Output
Plotted on a Linear
Scale for Ease in Com-
parison with Experi-
mental Data in Fig. 14

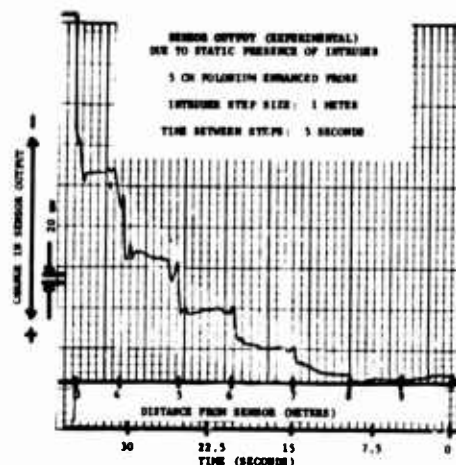


Fig. 17

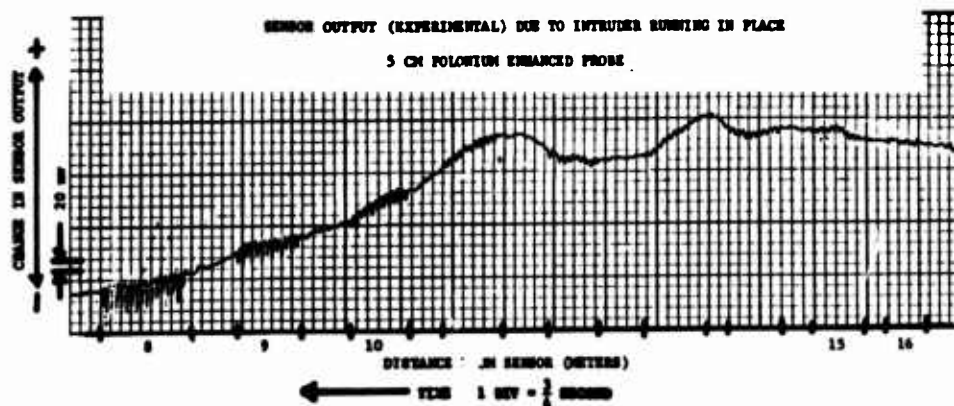


Fig. 18

A NEW MONOLITHIC CRYSTAL FILTER USING PIEZO-
ELECTRIC COUPLING BETWEEN ANISOTROPIC LAMINAE

ARTHUR BALLATO, PhD
THEODORE LUKASZEK, Mr.
USA Electronics Technology and Devices Laboratory (ECOM)
Fort Monmouth, New Jersey 07703

INTRODUCTION

Resonant crystal elements for frequency control and selection have been available for fifty years. They have come to play an indispensable role in modern communications as a result of a steady stream of improvements in manufacturing technology. During the past twenty years, sophisticated mathematical investigations have been carried out on the properties of mechanically vibrating piezoelectric crystal plates from which crystal resonators are constructed. Unfortunately, the mathematical results are particularly abstruse, and little progress has been made in this interval to bring the analytical results to bear on practical designs. On the other hand, new requirements for Army secure tactical communications make it appear highly unlikely that any further significant improvements in conventional crystal technology can be effected without some new basic contribution by theory.

Yet the problem of obtaining practically useful results from complex mathematical formulations is only part of a larger difficulty - - any new devices must meet severe Army stipulations of high reliability, small size and cost. In addition, it appears that conventional configurations of resonant elements may be inherently incapable of providing certain characteristics that can be envisioned as future requirements.

Our solution of this dilemma involves a novel stacked-resonator configuration that forms the basis for a new family of frequency-selective devices, plus a systematic method of analysis based upon

Preceding page blank

distributed network theory, whereby the physics of crystal-plate vibrations may be interpreted in a relatively uncomplicated manner. Devices formed from such stacked resonators can perform a wide variety of frequency- and time-control functions; in this paper, we will confine our attention solely to frequency-domain filters. The analysis and network representation procedure to be given here is even more general; it can be applied to surface as well as bulk modes of vibration, and to the interaction of such acoustic modes with electromagnetic fields.

Traditional crystal filters are realized by wiring individual crystal resonators into an appropriate network, the resonance interactions taking place via the electrical interconnections. Figure 1a shows schematically a conventional single-crystal resonator. Figure 1b shows a monolithic crystal filter employing acoustic coupling taking place laterally between two resonators on a single plate. This more recently introduced filter type is somewhat smaller than the conventional variety, but is more fragile and more limited in operating characteristics. The structure to be analyzed in this paper is given in Fig. 2; here the individual crystal plates are stacked up in the thickness direction. Mechanical and piezoelectric coupling take place at the interface between plates when they are driven in modes of motion that depend on the thickness coordinate only; this coupling can be large compared to lateral coupling, resulting in new ranges of characteristics. Moreover, the sandwich configuration is miniature and robust, and construction requires no advance over current technology.

To characterize a device of this nature, it is necessary to have a description of bulk acoustic-wave propagation in anisotropic media that are piezoelectric. Then, the piezoelectric driving mechanism must be treated, and finally, the joining of two or more laminae has to be described. After this has been accomplished, the behavior of such devices can be simulated on a computer, and optimized for a given application.

TRANSMISSION-LINE REPRESENTATION OF A PIEZOELECTRIC PLATE

Thickness modes depend only upon the thickness coordinate; a crystal plate supports three families of plane waves that progress in this direction, corresponding to one predominately thickness-stretch and two orthogonally polarized predominately thickness-shear modes. We wish to show that these plane wave modes obey Heaviside's transmission line equations (1) even when the piezoelectric effect is included.*

*Historically, Christoffel, in 1877, gave the equations describing

To accomplish this in the simplest fashion, we transform the components of elastic stress and displacement, and the piezoelectric constants, to the basis of the eigenvectors instead of using the usual laboratory coordinates.** In normal coordinates the elastic cross-coupling vanishes, and Newton's equation of motion is just

$$\dot{T}_{3m,3} = -\rho \omega^2 u_m^{\circ} \quad (m = 1, 2, 3) \quad [1]$$

Hooke's law similarly reduces to

$$\dot{T}_{3m} = \kappa_m u_{m,3}^{\circ} + e_{33m} a_3 \quad [2]$$

where T° , u° and e° are, respectively, the components of stress, displacement, and piezoelectric constant referred to the eigenvectors of the problem, and ρ , ω , κ and a are respectively the mass density, circular frequency, elastic eigenvalue, and a spatial constant representing electric field strength. The subscript "3" refers to the x_3 axis, taken in the thickness direction, a comma preceding a subscript 3 denotes $\partial/\partial x_3$, and a time dependence $\exp[+j\omega t]$ has been assumed.

Separating the stresses into "wavy" and constant terms,

$$\dot{T}_{3m} = \tilde{T}_{3m}^{\circ} + \bar{T}_{3m}^{\circ} \quad [3]$$

allows [1] and [2] to be written as

$$\tilde{T}_{3m,3}^{\circ} = -\rho \omega^2 u_m^{\circ} \quad , \quad \text{and} \quad \kappa_m u_{m,3}^{\circ} = \tilde{T}_{3m}^{\circ} \quad [4], [5]$$

The transmission-line modal voltages, V_m , and currents, I_m , are identified with the mechanical quantities as follows:

$$V_m = A \tilde{T}_{3m}^{\circ} \quad \text{and} \quad I_m = -j\omega u_m^{\circ} \quad [6], [7]$$

where A is the area perpendicular to the thickness. Then, with the further definitions $Z_m = A \sqrt{\rho c_m}$, $Y_m = Z_m^{-1}$, and $x_m = \omega \sqrt{\rho/c_m}$, it is seen that three acoustic plane waves each obey the Heaviside equations (1):

plane acoustic wave propagation in crystals of arbitrary anisotropy(2). In 1941, Lawson generalized the treatment to piezoelectric crystals (3). The exact solution for a piezoelectric crystal plate driven by a thickness-directed electric field in thickness modes of vibration was then obtained by Tiersen in 1963 (4).

**This was first done for this problem in 1970 by Yamada & Nizeki (5).

$$V_{m,3} = -j \chi_m Z_m I_m; I_{m,3} = -j \chi_m Y_m V_m. \quad [8], [9]$$

From this it follows that the distributed network representation of a thickness mode plate consists of three modal transmission lines to "carry" the piezoelectric waves. At the boundaries of the plate, however, satisfaction of the stress boundary conditions requires that the constant term \bar{T}_{3m}^* from [3], which is equal to $e_{33m}^0 a_3$, be accounted for. Physically, it means that the modes are coupled, at the boundaries, by the piezoelectric effect; from the network standpoint it amounts to placing the piezo drive transformers at the plate surfaces. We show this in Fig. 3 for the case where it is assumed that a single mode only is driven. The piezo transformer turns ratios are

$$n_m = A e_{33m}^0 / (2h), \quad [10]$$

where $2h$ is the plate thickness. In the figure each transmission line carries one mode with a velocity $v_m = \omega / \chi_m$. The static capacitance C_0 , between the electrodes, equals $A \epsilon_{33}^0 / (2h)$. The inertial effect of the electrode masses is represented by the lumped inductances at the transmission-line ends; these are in series (for the driven mode shown) with the piezoelectric transformer secondaries that exert surface tractions \bar{T}_{3m}^* on the plate to drive it when a voltage is impressed across the electrodes. The negative C_0 shown leading from the transformers is a consequence of the self-electric-field produced by the piezoelectric waves as they propagate. This self-field opposes that applied via the electrodes.

Because the electrode inertia in Fig. 3 is isotropic, the three inductances at each surface are equal, and no transformation of the boundary impedance is necessary when viewing the system from either the normal- or laboratory-coordinate framework. In any other case it is necessary to add, at each surface, a transformation network to convert between the coordinate systems. This has been done in Fig. 4, which gives the complete and exact result for the problem of a single plate driven in thickness excitation of thickness modes (TE_{TM}). At first sight, this representation might appear as intricate as the mathematical analysis it is intended to supplant, but one can see that each portion of the schematic has a definite physical meaning, and that each of the relevant physical effects has been separated out. Furthermore, we are now in a position to apply results from network theory for both analysis and synthesis, and, equally importantly, we may now treat the network of each plate as a simple building-block in obtaining results for multi-layer stacks. In the next section we will consider the effect of mechanically stacking laminae of crystals.

INTERFACE MECHANICAL COUPLING

The transformation between the eigenvector and laboratory coordinate systems is orthogonal, and its network realization is given in Fig. 5. The network is composed of ideal transformers with ratios, β_{im} , that are the components of the eigenvectors of each mode m . In the solution to the plane-wave propagation problem these quantities appear in a natural fashion upon solving for the three elastic stiffnesses c_m . They depend upon the elasto-piezo-dielectric matrix of the particular material used and the orientation of the plate thickness with respect to the crystallographic axes. At the ports representing the x_i coordinates, the network provides directly the untransformed stresses and displacements, T and u , required for satisfaction of the plate boundary conditions. For the practically important case of two crystals of arbitrary orientation in welded mechanical contact, the three components each of stress and displacement are continuous across the interface. This requires only that the two boundary networks (one for each plate), of the type of Fig. 5, be connected port-to-port. When the resulting primary circuit loops are suppressed, the complete network appears as in Fig. 6.

Figure 6 is the most general case of mechanical interfacial coupling; in most instances it will simplify considerably, for either of two reasons. Firstly, the crystal material properties and/or the orientation direction of one or both of the plates may lead to a set of eigenvectors where some of the β_{im} vanish; secondly, boundary conditions other than welded may be prescribed, as, for example, would be the case for two surfaces with lateral slip.

Piezoelectric effects have been omitted from Fig. 6 for simplicity. In practice they are merely added by placing in series with each of the transmission lines a piezo transformer of appropriate turns ratio, with the primary connected to the electrical port as in Fig. 4.

NETWORKS FOR CRYSTAL LAMINAE

Having reached a network interpretation for the thickness modes in an arbitrarily anisotropic and piezoelectric plate, and having described in network terms the joining of two such plates, we now turn to the use of these circuits for implementing practical device designs. In order not to obscure the approach, we take a simple, specialized case of our general networks and proceed to a discussion of frequency-domain filters. It is clear that insofar as Fig. 4 completely represents a single plate, it may be used as a building-block to represent a laminae of many plates merely by a tandem interconnection of networks of this sort. The simplest situation is the layering of two lamina.

Even this case appears rather complex when all piezoelectric constants are present, along with general mechanical coupling; however, when a single mode in each plate is piezoelectrically driven, and when the materials and/or crystal orientations are arranged so that the boundary network simplifies, the two-layer stacked-crystal filter that emerges is readily analyzed. Figure 7 shows this case. The input and output, seen in Fig. 2, correspond to terminal pairs A, B and C, D. These ports provide electrical entry to the piezoelectric remainder of the network via the piezo transformers located at the surfaces of each plate. The lumped inductances, as in Fig. 3, represent the electrode masses, while the shorter transmission line interposed between the lamina describes the influence of the bonding layer that may be, e.g., indium for VHF frequencies, or even commercial epoxy cements at lower frequencies. The coupling layer is often not of negligible thickness and must be described by a distributed network component. It is usually not sufficient to regard it as lossless either, in which case the effects of viscosity are incorporated by making the transmission line parameters complex.

Figure 8 depicts a two-layer stack, in each plate of which a single mode is piezoelectrically active. The outer surfaces of the stack are assumed coated with electrodes of negligible thickness, and appear in the figure as short circuits across the transmission lines; the coupling layer is likewise neglected, but now the interfacial network representing the welded contact is included. The left-hand portion of the boundary network is somewhat simpler than the completely general right-hand side. This particular boundary network describes the situation where an elastically monoclinic crystal is joined to a triclinic crystal, with the reference coordinates taken to coincide with the monoclinic crystal axes. A quartz plate rotated about the x_1 axis has the class 32 symmetry of quartz destroyed by the rotation and appears as a crystal of the monoclinic class 2. If the two plates of the stack are further rotated with respect to each other, about the common thickness axis, as shown in Fig. 2 by the non-parallel edges, the monoclinic plate will then appear in the reference coordinates to lack any symmetry and we must use all nine β_{im} in the boundary circuit. Thus Fig. 8 represents two rotated-Y-cut quartz plates in welded contact. A computer simulation of the attenuation function versus frequency, normalized to the fundamental resonance of a single plate, is given in Fig. 9. The filter device is assumed to be operating in a 50 ohm system, and the electrode area is arranged so that the reactance of C_0 is -50 ohms at the fundamental plate resonance. Both plates are assumed to have the same resonance frequency, which is 100 MHz, and to be AT-cuts. The angle of mutual rotation is ψ ; by varying it, the turns ratios β_{im} are altered, along with the filter response.

EXPERIMENTAL REALIZATIONS OF STACKED-CRYSTAL FILTERS

In this section we discuss experimental results based on the foregoing theory. Figure 10 corresponds to Fig. 2. It shows two AT-cut quartz plates, joined as a stacked-crystal filter device, and mounted in a flatpack. The crystals are circular (14 mm diameter); future designs will be further miniaturized. The potentiality of these filters is such that they can be made completely compatible with micro-circuitry. For experimental purposes, square plates are easier to align for desired values of Ψ . Two square, 25 mm, AT-cut plates with individual resonances at 2.16 MHz, bonded with an epoxy composition 2 μ m thick, were used to produce the filter characteristics of Fig. 11. A simple tuning network was used at both ends of the filter to provide some matching to the remaining circuitry. By this means a reactive balance was obtained; the resistive unbalance accounts for 6 dB of the total insertion loss of about 8 dB, and can be eliminated by further matching. It appears reasonable that the total in-band loss can be brought well under 2 dB. This unit has a fractional bandwidth at the 10 dB point of 4.2%. A capacitive coupling between resonators accounts for the finite pole of attenuation located 5% below the center frequency. The two plates have a mutual rotation angle of 80° ; this produces the abrupt 5 dB jump in attenuation occurring 14% above the center frequency, because the driven shear mode couples acoustic energy into the piezoelectrically-inactive shear mode of both plates via the boundary transformer network, as seen in Fig. 8.

By moving the two shear modes closer together, multi-mode interactions can be produced, leading to new wide-band filter capabilities. One accomplishes this movement by changing the angle at which one or both of the crystals is cut. Figure 12 is a sample of what may be obtained. Plotted are the three velocity ratios that result for rotated Y-cut-quartz plates. For $\theta = +35\frac{1}{4}^\circ$, the AT-cut angle, the ratio, σ_b/σ_c , of the two shear modes, is 1.14; the separation decreases for a range of negative angles, which can be used for wide-band filters. Similar results are arrived at using highly piezoelectric crystals such as lithium niobate and lithium tantalate. Calculations show that both of these substances are richly supplied with suitable orientations for multi-mode designs.

When two or more stacked-crystal filters are arranged electrically in cascade, the result is an improvement in the discrimination. Figure 13 gives the response of two stacked-filters in cascade. Each has a response similar to that of Fig. 11; together, the shape factor has been improved significantly to approximately 2 to 1, with a 10 dB bandwidth of 3.1%.

A variety of 10-MHz stacked-crystal experimental filters have been constructed using circular plates. These can be made so that capacitive coupling is extremely small and a symmetrical response results, as shown in Fig. 14. The relative angle ψ is zero, so the circuit of Fig. 7 applies. A simple tuning network at both ends has been used here, as in earlier work. With these 10 MHz designs, as with the others, a most encouraging side benefit has been observed. One of the greatest problems with conventional crystals for filter application is the presence of unwanted modes of vibration that appear in the filter frequency spectrum as undesired passbands in the reject region and rejection regions in the passband. The stacked-crystal filter device appears to be extremely free from such disturbances, due to the clamping that takes place at the joined surfaces. In proof of this, a typical sweep of a 10-MHz filter unit is shown in Fig. 15; from 1 to over 100 MHz all unwanted mode activity is suppressed by at least 50 dB.

The above theory and experimental results have concentrated on two-layer laminae; yet the theory extends straightforwardly to any number of crystals. A sampling of devices proposed for future investigations is shown in Fig. 16. The multi-layer, multi-mode configurations have possible applications as frequency- and time-domain filters, delay lines, pulse and code generators, and frequency discriminators. For each crystal in the stack, the choice of material, thickness, and orientation, plus the mutual rotation angles between plates, offers wide choices of parameters for the satisfaction of practical design characteristics.

For the single case of bandpass filters, specific applications of these devices are to output filtering, and front-end filtering in proximity to transmitters operating in adjacent bands, where moderate power levels must be handled with a minimum of modulation and distortion effects. Nonlinear effects at levels that are significant in these applications, because of the production of intermodulation, for example, are generated in most solids at acoustic power flow densities of less than one watt per square millimeter, so bulk waves, as utilized in this treatment, are to be favored for these uses.

SUMMARY & CONCLUSION

This paper presents a systematic and comprehensive analytical treatment of thickness-mode plate vibrations in stacked piezoelectric crystals that is at once rigorous and practical. The results are cast in terms of distributed electrical networks that reduce arduous mathematical procedures to standard circuit theory and network manipulations. These are then applied to analyze a novel type of frequency-sensitive device - the stacked-crystal filter. Experimental realizations of

this device prove what is predicted on the basis of computer simulations applied to the networks: that this family of devices, and, by extension, other classes of devices based upon the stacking principle applied to piezoelectric crystal laminae, possess a striking potential for satisfying current and future frequency- and time-control requirements of Army secure communications. In particular, highly temperature-stable stacked-crystal bandpass filters are inherently capable of wideband operation with low insertion loss and extraordinary mode spectral purity.

REFERENCES

1. W. L. Weeks, Electromagnetic Theory for Engineering Applications, (Wiley, New York, 1964), pp. 62-77.
2. E. B. Christoffel, "Ueber die Fortpflanzung von Stossen durch elastische feste Korper," *Annali di Matematica (Milano)*, Series II, Vol. VIII, 1877, pp. 193-243.
3. A. W. Lawson, "Comment on the Elastic Constants of Alpha-Quartz," *Phys. Rev.*, Vol. 59, 1941, pp. 838-839.
4. H. F. Tiersten, "Thickness Vibrations of Piezoelectric Plates," *J. Acoust. Soc. Amer.*, Vol. 35, No. 1, Jan. 1963, pp. 53-58.
5. T. Yamada and N. Niizeki, "Admittance of Piezoelectric Plates Vibrating Under the Perpendicular Field Excitation," *Proc. IEEE* Vol. 58, No. 6, June 1970, pp. 941-942.

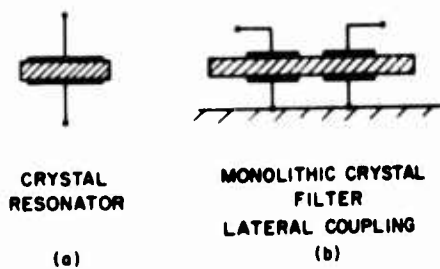


Fig. 1 Traditional frequency-selective devices.



Fig. 2 Novel stacked-crystal filter device.

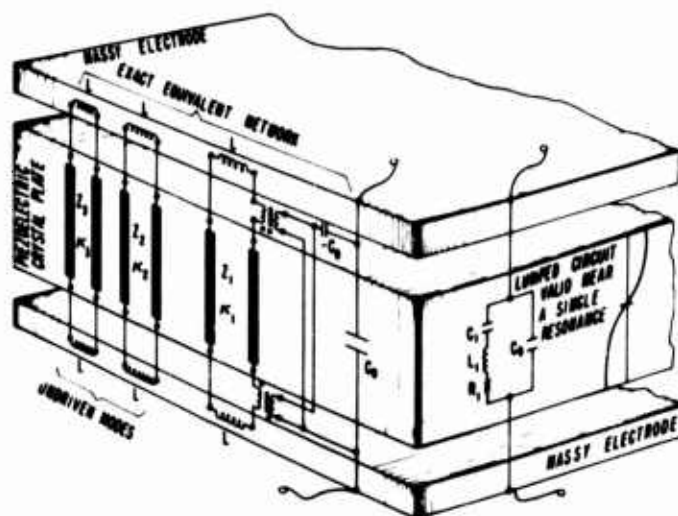


Fig. 3 EXPLODED VIEW OF A CRYSTAL PLATE WITH MASSY ELECTRODES, FOR THE CASE OF A SINGLE THICKNESS MODE PIEZOELECTRICALLY DRIVEN
SUPERPOSED ARE THE EXACT AND LUMPED APPROXIMATE EQUIVALENT ELECTRICAL NETWORKS

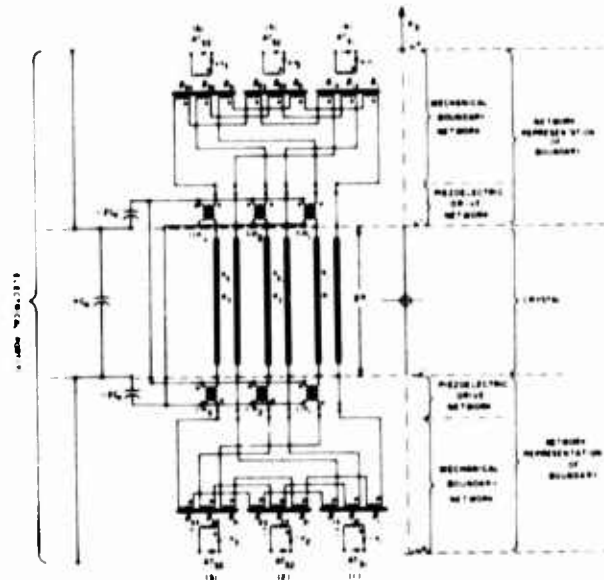


Fig. 4 EXACT ANALOG REPRESENTATION OF TEM PLATE, FOR ARBITRARY BOUNDARY CONDITIONS

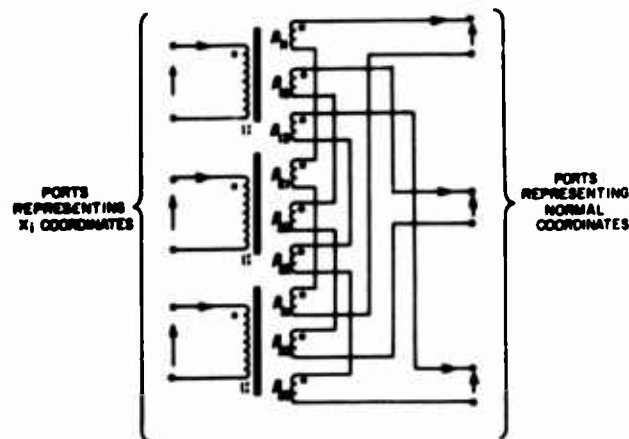


Fig. 5 IDEAL TRANSFORMER REALIZATION OF AN ORTHOGONAL TRANSFORMATION: THE MECHANICAL BOUNDARY NETWORK.

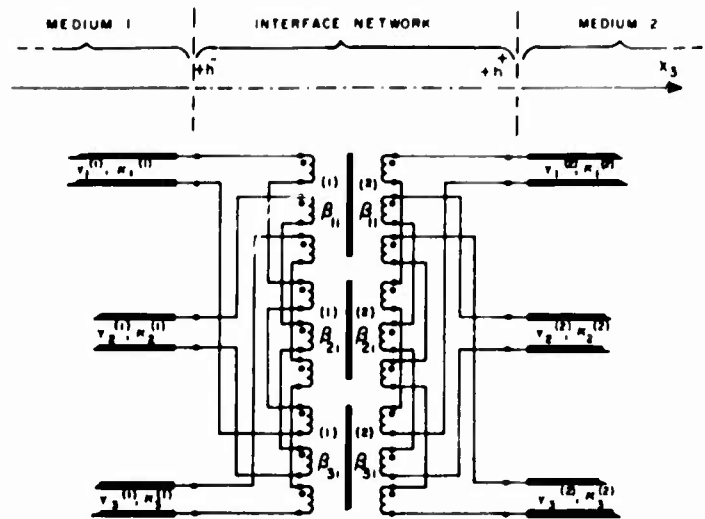


Fig. 6 EXACT REPRESENTATION OF MECHANICAL COUPLING AT THE INTERFACE OF TWO ARBITRARILY ANISOTROPIC MEDIA. PLANE ACOUSTIC WAVE PROPAGATION NORMAL TO BOUNDARY. PIEZOELECTRIC DRIVE TRANSFORMERS OMITTED FOR CLARITY.

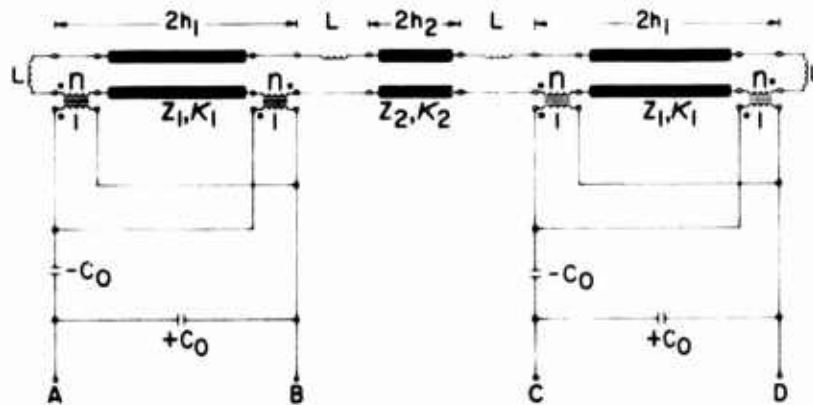


Fig. 7 Two-layer, single-mode network representation including distributed coupling layer and massy electrodes.

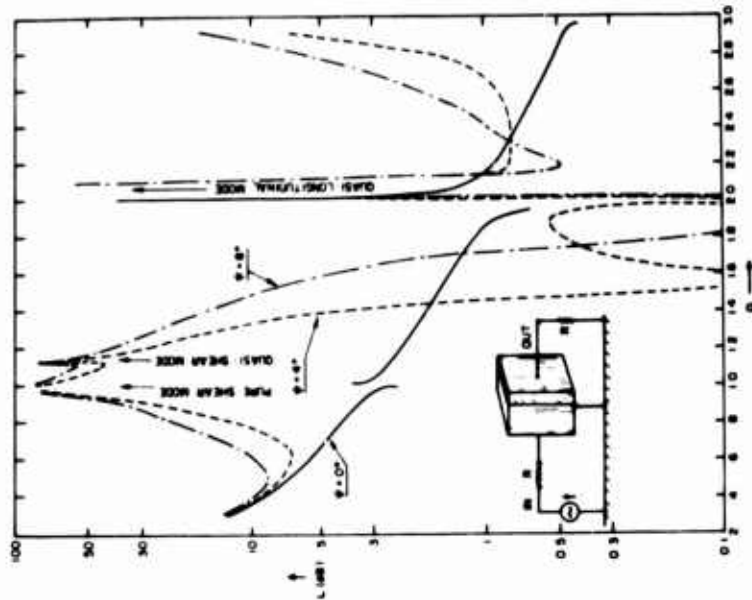


Fig. 9 ATTENUATION VERSUS FREQUENCY FOR A TWO-LAYER STACK OF 100 MHz AT-CUT RESONATORS AS FUNCTION OF RELATIVE ROTATION ABOUT COMMON THICKNESS AXIS

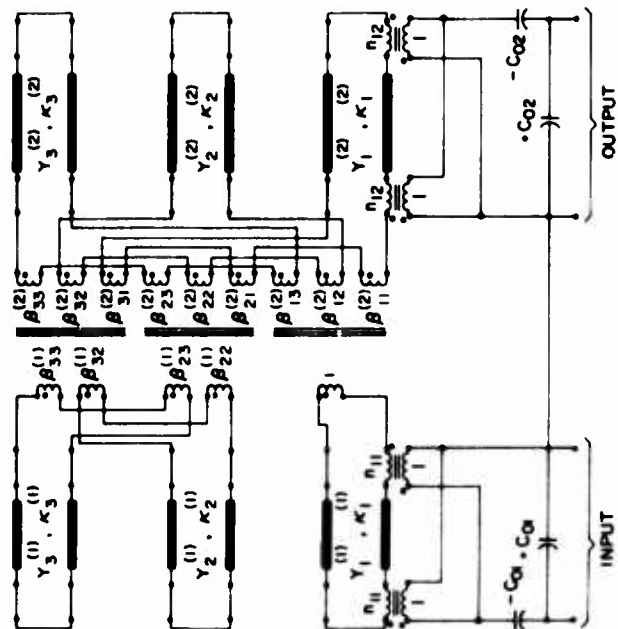


Fig. 8 Network description of two bonded, AT-cut resonators with arbitrary mutual rotation.



Fig. 11 Response of two-layer stack.

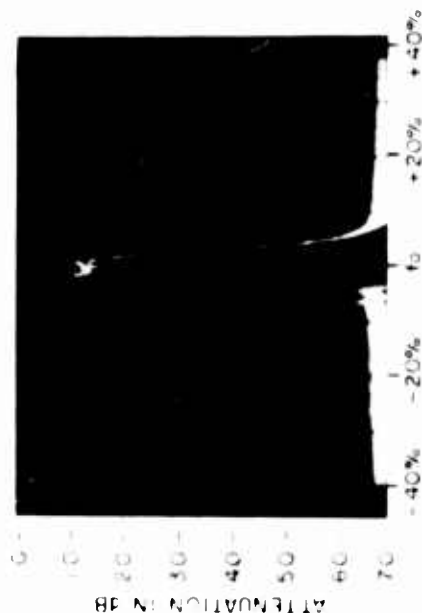


Fig. 13 Cascade of two stacked devices.

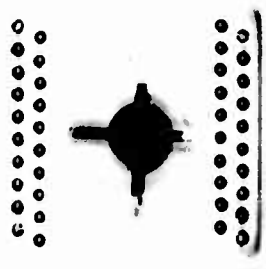


Fig. 10 Experimental filter unit in flatpack.

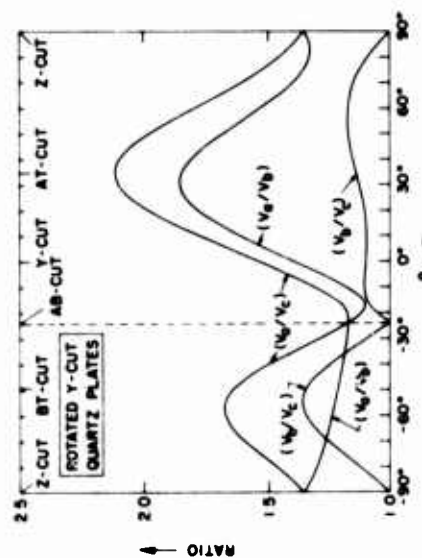


Fig. 12 Modal velocities vs orientation angle.

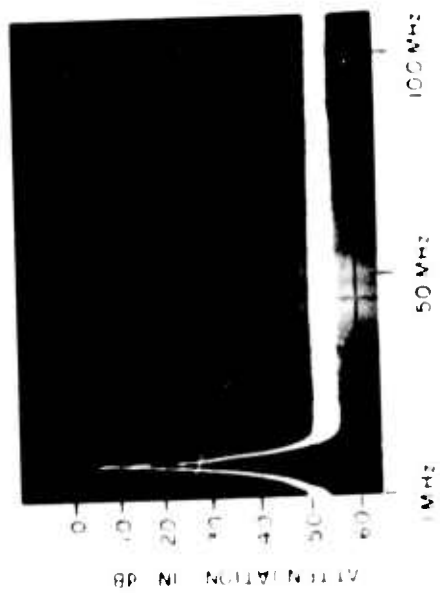


Fig. 14 Symmetrical bandpass response.

Fig. 15 Demonstration of spectral purity

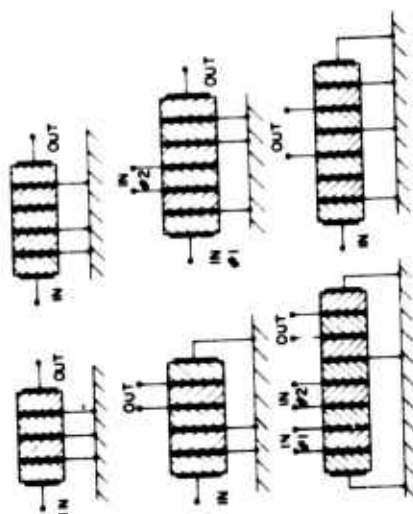


Fig. 16 Proposed multi-layer stacks.

NUMERICAL SIMULATION OF TRANSONIC
FLOW ABOUT AIRPLANES AND HELICOPTER ROTORS

DR. V. F. BALLHAUS
F. X. CARADONNA
AMES DIRECTORATE, USAAMRDL
MOFFETT FIELD, CA 94035

INTRODUCTION

Most modern aircraft achieve optimum cruise performance and maneuverability when flying at high subsonic Mach numbers in the transonic regime. A typical flowfield about an airfoil operating in this range is shown in Fig. 1. The flow is everywhere subsonic except in embedded supersonic regions which are usually terminated by drag-producing shock waves. As the Mach number increases beyond the optimum, the extent and strength of these shock waves increases, and performance deteriorates rapidly giving way to flow separation and buffet.

The onset of such adverse transonic effects also severely limits the performance of helicopter rotors. To avoid excessive strengthening of embedded shock waves, maximum rotor tip Mach numbers must be less than about $M_t = 0.85$. Hence, any increase in forward flight speed must be compensated by a decrease in rotor RPM. Both of these changes in speed decrease the dynamic head available for lift on the retreating side. To avoid a roll moment, the lift produced on the advancing and retreating sides must be about equal.

This requires an increase in retreating blade angle of attack. Eventually a point is reached at which an increase in forward flight speed cannot be balanced by an increase in angle of attack on the retreating side without encountering retreating-blade stall. At this point, an increase in forward flight speed can only be effected by an improved transonic tip design which would delay the onset of adverse transonic effects to some larger tip Mach number, $M_t = 0.92$ for example.

Preceding page blank

Before attempting any such improvement in design, one should determine what physical parameters are of primary importance in transonic flow and how to include them in an analysis. First, transonic flows, because of their mixed subsonic-supersonic flow character, and because of the presence of embedded shock waves, are inherently non-linear. Thus, for example, the effects of airfoil thickness and camber cannot be treated independently and then superposed to obtain a composite solution as in linear theory. Second, because of the strong lateral influence of disturbances in transonic flows, small variations in geometry can have significant effects on the overall flowfield. For this reason, most practical transonic flow problems for both airplanes and helicopter rotors are inherently three-dimensional, and performance is highly configuration dependent. For the same reason, the effects of viscosity -- i.e., the presence of the boundary layer, wake, and shock-boundary layer interaction -- can significantly affect the overall flowfield even for high Reynolds numbers. Finally, transonic flows are unsteady in the sense that the effect of a perturbation in transonic flow conditions dies out much more slowly than in a purely supersonic or subsonic flow. This is because unsteady disturbances travel upstream at a speed nearly equal to that of the wing instead of propagating rapidly away from it.

Although there has been a considerable effort over the past two decades to compute transonic flowfields analytically and to simulate them experimentally, predictions of in-flight performance for new configurations are often unreliable. The analytical approach has failed primarily because the equations governing a transonic flowfield are non-linear, and there exists no general method of solution. Experimental simulations, or wind tunnel tests, have three major weaknesses. First, because of the strong lateral influence present in transonic flows, wind tunnel walls often significantly affect test results. Second, it is difficult to properly account for the effects of viscosity because of the large difference between Reynolds numbers currently attainable in wind tunnels and those encountered in flight. Third, and most important, is cost.

The cost, measured in total wind tunnel test hours, for development of a new aircraft has been increasing at nearly an exponential rate over the past two decades. For example, it is estimated that the B1 bomber will require 30,000 hours of wind tunnel testing at a cost of about \$30 million. Fully instrumented models cost on the order of \$300 thousand and transonic tests cost about \$2,000 per hour. So, experimental parametric studies, i.e., varying geometry to optimize performance, must be held to a minimum. Hence, designers must rely heavily on their experience. This usually precludes consideration of configurations that are radically different from those that have been

built previously. The testing of transonic rotor tips is even more difficult and expensive because of the rotary environment. In fact, to our knowledge, there exist no reliable surface pressure data for transonic flow about rotors. The only design tools available are experience and two-dimensional strip theory.

On the other hand, computer technology has been improving at nearly an exponential rate. Computer speed has increased by a factor of ten every three years, while the cost of computing a given calculation has decreased by a factor of ten every five years. It seems reasonable, then, to attempt to use numerical simulations to supplement wind tunnel tests whenever possible. Three years ago an efficient relaxation algorithm was developed to simulate steady, inviscid, transonic flows about airfoils (ref. 1). Computed results compared with a wide variety of experimental data have established the usefulness of the method for predicting two-dimensional, transonic flows. The authors and their associates have subsequently extended the method to treat wings, wing-cylinder combinations, and helicopter rotors (ref. 2-6).

THE METHOD

The governing equation is the transonic, small-disturbance, velocity-potential equation

$$(1-M_\infty^2 f^2(y) - (\gamma+1) M_\infty^2 f(y) \phi_x) \phi_{xx} + \phi_{yy} + \phi_{zz} = 0$$

where

$f(y) = y/R$ (ratio of span distance to blade radius) for a rotor,

$= 1$ for a fixed wing,

M_∞ = freestream Mach number,

γ = ratio of specific heats, and

x, y, z are defined in Fig. 2.

The equation is elliptic when the coefficient of ϕ_{xx} is positive, corresponding to subsonic flow, and hyperbolic when the coefficient of ϕ_{xx} is negative, corresponding to supersonic flow.

The derivatives ϕ_{xx} , ϕ_{yy} , ϕ_{zz} are approximated by finite differences, where central differences are used for ϕ_{yy} and ϕ_{zz} . For stability, central or backward differences are used for ϕ_{xx} in subsonic and supersonic flow regions, respectively (see Fig. 3). This switching from backward to central differences allows for the

automatic capture of embedded shock waves. The difference equations provide a matrix equation consisting of N nonlinear algebraic equations in N unknowns, where N is the number of mesh points at which ϕ is to be evaluated. Typical values of N are 50,000 for helicopter rotors and 200,000 for fixed wing aircraft. The matrix equation is solved by first guessing the solution vector ϕ and then iteratively refining ϕ until the system of equations is satisfied to some predetermined order. The iteration procedure used is successive line over-relaxation.

The primary difficulty associated with three-dimensional flow simulations is the proper treatment of complex boundary conditions. The most effective approach has been to transform swept, tapered planforms into rectangles at the beginning of the solution process. This transfers the effects of nonsimple geometry from the boundary conditions to the equations of motion thereby complicating the difference operators and stability criterion.

The ultimate objective is to develop a computer simulation code that can be used for the design and analysis of airplanes and high speed helicopter rotor tips. Our approach has been to treat successively the parameters that are of primary importance in transonic flows -- nonlinearity, three-dimensionality and, ultimately, unsteadiness and viscosity. As the program is generalized, computed results are compared with experimental data for fixed-wing aircraft. The program is then used to simulate flows about rotary wings, for which no experimental data is available. At this point, the code is capable of simulating steady, non-linear, three-dimensional flows. An unsteady, nonlinear, three-dimensional code is currently under development. Presently the effects of viscosity are accounted for only as empirical corrections to the inviscid results. The complete treatment of viscosity awaits the introduction of efficient turbulence modelling techniques.

RESULTS

The first correlation to verify the computed results was with experimental data for the C-141 airplane wing (see Fig. 4). The C-141 was chosen because of its general configuration -- i.e., it had sweep, taper, and twist -- and because there was a variety of both wind tunnel and flight data available. Before comparing, it is helpful to analyze the experimental data at various Mach numbers and Reynolds numbers to determine under what conditions good agreement between computed and experimental results might be expected.

Upper surface pressure coefficients measured in the wind tunnel and in flight are shown in Fig. 5. A comparison of results at $M_\infty = 0.825$ for the low-Reynolds-number (8.5×10^6) wind tunnel case and high-Reynolds-number (36×10^6) flight case shows a discrepancy in shock wave location of about 15%. This was a well-known case of unreliable wind tunnel prediction of in-flight conditions, and it resulted in a 100% error in pitching moment. The discrepancy is due to a loss in lift in the wind tunnel case resulting from trailing edge separation, which is Reynolds number dependent. Therefore, one might expect the computed results, which do not account for viscosity and, hence, represent infinite Reynolds number unseparated flows, to agree better with the flight results than with the wind tunnel.

A comparison of high-Reynolds-number flight results at Mach numbers of 0.825 and 0.85 again shows a discrepancy in shock location. This is caused by shock-induced separation in the $M_\infty = 0.85$ case. Shock-induced separation occurs at high Mach numbers where the embedded shock has strengthened to the point that it separates the boundary layer. Although the inviscid results cannot properly simulate the flowfield once shock induced separation has occurred, they can be used to predict with considerable accuracy at what Mach numbers it will occur (see ref 6). To summarize, then, computed results should compare favorably with experiment when Reynolds number is sufficiently high and Mach number sufficiently low that the flow remains unseparated. Computed and experimental surface pressures are compared in Fig. 6 for $M_\infty = 0.825$. The computed pressures agree much better with the high-Reynolds-number, unseparated flight data than with the low-Reynolds-number, wind tunnel results. It should be mentioned that, even though typical helicopter Reynolds numbers are of the same order as those reached in wind tunnels, advancing rotors operate at nearly zero lift in the transonic regime, so trailing edge separation would not have as large an effect on shock location.

Once the accuracy of the numerical simulation had been established for fixed-wing aircraft, it was generalized to treat rotary wings. The model chosen was the simplest that maintained the salient features of high speed tip flow. The transonic small disturbance equation for the hover condition was solved for a nonlifting rotor tip of arbitrary planform and profile. The most significant simplification is the neglecting of viscous effects and unsteadiness in incident Mach number. Included in the model are the effects of rotation, nonlinear mixed flow, spanwise freestream Mach number gradient, and flow relief resulting from tip geometry.

A series of rotor configurations was calculated starting with the usual rectangular blades having NACA-0012 profiles. Fig. 7 shows the

pressure distribution on one such blade. A good quantity by which to assess the significance of such a calculation is the section drag. Fig. 8 shows a comparison of section drag for three rectangular blades of different aspect ratio along with the equivalent two-dimensional section drags ($AR = \infty$). This graph is a dramatic indicator of how bad an estimate of drag is provided by two-dimensional strip theory. In these cases, no significant portion of the blade flow is accurately predicted by two-dimensional results. Furthermore, it is seen that in the range of practical aspect ratios, wave drag is approximately proportional to aspect ratio. This is not surprising in view of the strong lateral signal propagation characteristic of transonic flow.

One can take greater advantage of the inherent three-dimensionality of transonic rotor flow by varying the tip planform. The results of calculations on simple swept tips are shown in the following. Fig. 9 shows the flow about a swept tip with two different section profiles, the NACA-0012 and a 12% circular arc. A comparison of Fig. 7 and Fig. 9a shows that sweep has effected a considerable improvement in flow. The shock strength has decreased markedly over most of the blade except for a small region at the tip where it has in fact increased. That this does not happen with the circular arc is a strong indication that profile cannot be considered independently of planform. The relief effect at the tip of a swept wing causes unsweeping, a shifting forward of the isobars. This implies that the leading edge expansion and subsequent compression will be accentuated. In the case of a NACA-0012, the combination of this with an already strong nose expansion results in an uncommonly great over-expansion and shock. The remedy for this would appear to be a reduction in the leading edge bluntness at the tip. To check this reasoning, an additional swept tip of identical planform was computed. However, in this case the profile used is a combination of the NACA-0012 and circular arc thickness distributions, given by

$$t(x) = \delta \frac{st_{0012}(X) + (1-S) t_{ca}(X)}{st_{0012}(XM) + (1-S) t_{ca}(XM)}$$

where δ = thickness ratio

t_{0012} = thickness distribution of a NACA-0012

t_{ca} = thickness distribution of a circular arc profile

XM = maximum thickness location

S = the percent of the thickness distribution given by the NACA-0012 distribution.

S varies linearly from 1.0 at the planform kink to .2 at the tip. Fig. 10 shows that this has been effective in removing the strong tip shock. The section drags for these configurations are compared in Fig. 11. (The reader should not be alarmed by the thrust at the tip; this merely reflects the unsweeping of isobars and, hence, the accentuation of leading edge expansion. The reverse effect occurs at the planform kink.) It is evident from the figure that the "mixed" profile out performs both the circular arc and the NACA-0012 profiles from which it is derived.

The lesson here is that planform variation can be very useful, but to be most beneficial, the profile used must be tailored to fit the planform. In the case of a swept tip, one must be cautious that the tip profile not have excessive leading edge expansion. This is an example of the process which might be followed in designing a rotor tip. To reduce costs, such computer parametric studies are first performed to select candidate designs before commitment to expensive fabrication and tests.

CONCLUDING REMARKS

Production computer codes have been developed which simulate three-dimensional transonic flows about fixed-wing aircraft and helicopter rotors. The fixed-wing code is presently being integrated into the design and analysis programs of aircraft companies and government agencies. For example, it is being used to perform parametric studies, at a cost of about \$150 per run, to select optimum light-weight fighter configurations for wind tunnel tests. It is also being used to modify the configurations of existing aircraft to improve their performance. The helicopter code is being used to design an improved transonic rotor tip to be built and tested later this year. Both codes are being reprogrammed to run on the new ILLIAC IV parallel processing computer which should reduce run costs by an order of magnitude. But even now the numerical simulations provide the designer with a relatively inexpensive analysis tool allowing him to consider a much wider range of design configurations before testing in the wind tunnel.

BALLHAUS & CARADONNA

REFERENCES

1. Murman, E. M.; and Cole, J. D.: Calculation of Plane Steady Transonic Flows. AIAA J., vol. 9, no. 1, Jan. 1971, pp. 114-121.
2. Caradonna, F. X.; and Isom, M. P.: Subsonic and Transonic Potential Flow Over Helicopter Rotor Blades. AIAA J., vol. 10, no. 12, Dec. 1972, pp. 1606-1612.
3. Ballhaus, W. F.; and Bailey, F. R.: Numerical Calculation of Transonic Flow About Swept Wings. AIAA Paper 72-677, June 1972.
4. Bailey, F. R.; and Ballhaus, W. F.: Relaxation Methods for Transonic Flow About Wing-Cylinder Combinations and Lifting Swept Wings. Lecture Notes in Physics, vol. 19. Proceedings of the Third Intl. Conf. on Numerical Methods in Fluid Dynamics, Paris, France, July 3-7, 1972, (H. Cabannes and R. Temam, eds.), Springer-Verlag.
5. Ballhaus, W. F.; and Caradonna, F. X.: The Effect of Planform Shape on the Transonic Flow Past Rotor Tips, AGARD Proceedings No. 111, Aerodynamics of Rotary Wings, Feb. 1973.
6. Lomax, Harvard; Bailey, Frank R.; and Ballhaus, William F.: On the Numerical Simulation of Three Dimensional Transonic Flow with Application to the C-141 Wing. NASA TN D-6933, Aug. 1973.

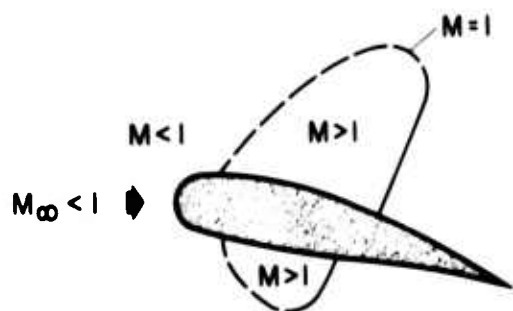


Fig. 1 Transonic flow

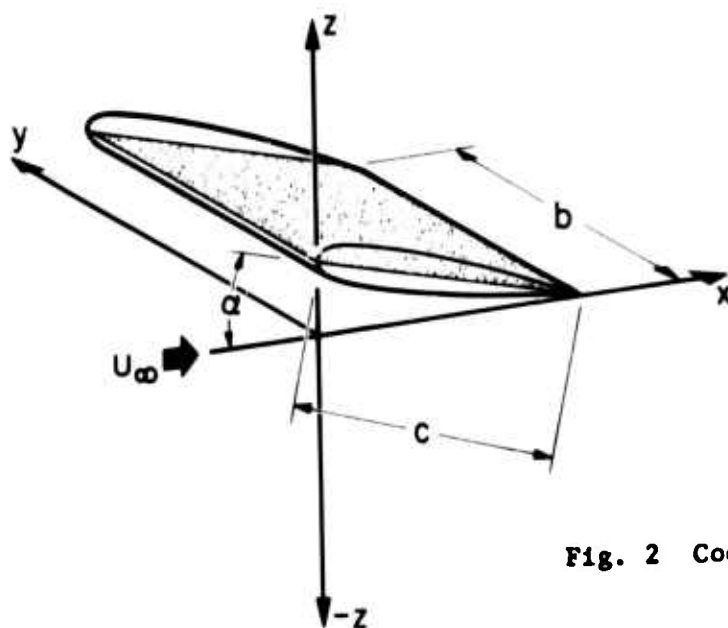


Fig. 2 Coordinates

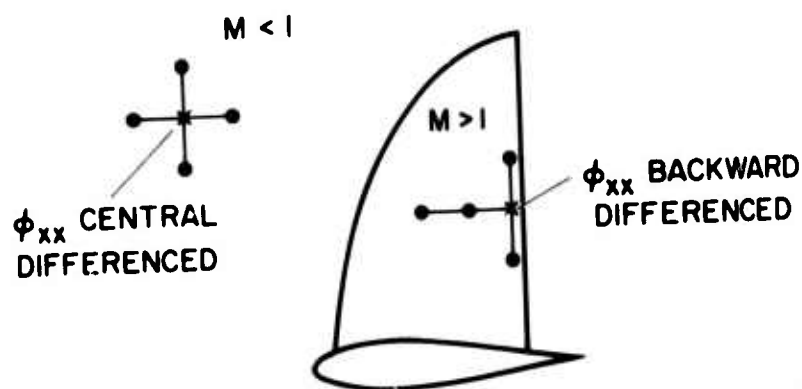


Fig. 3 The differencing scheme

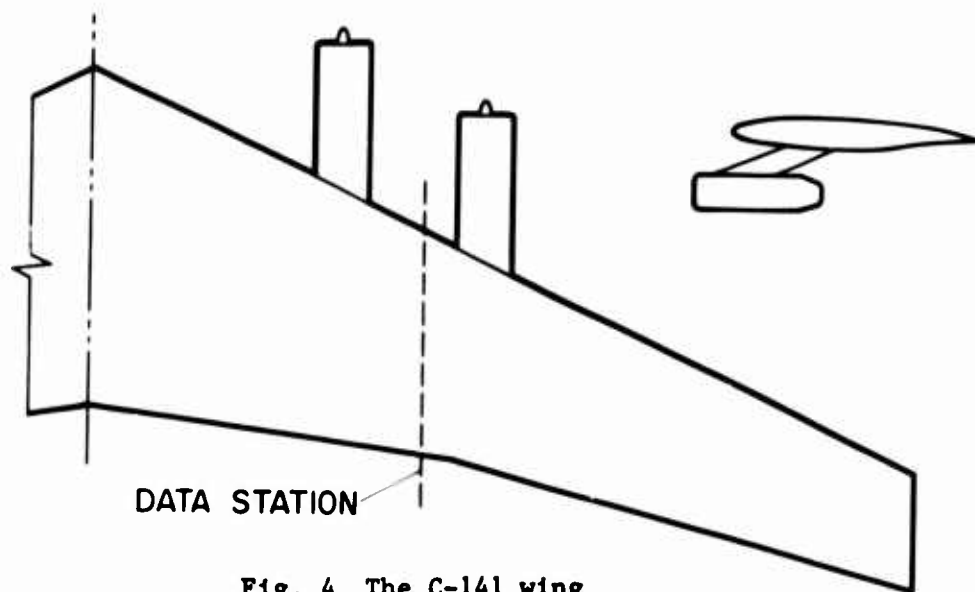


Fig. 4 The C-141 wing

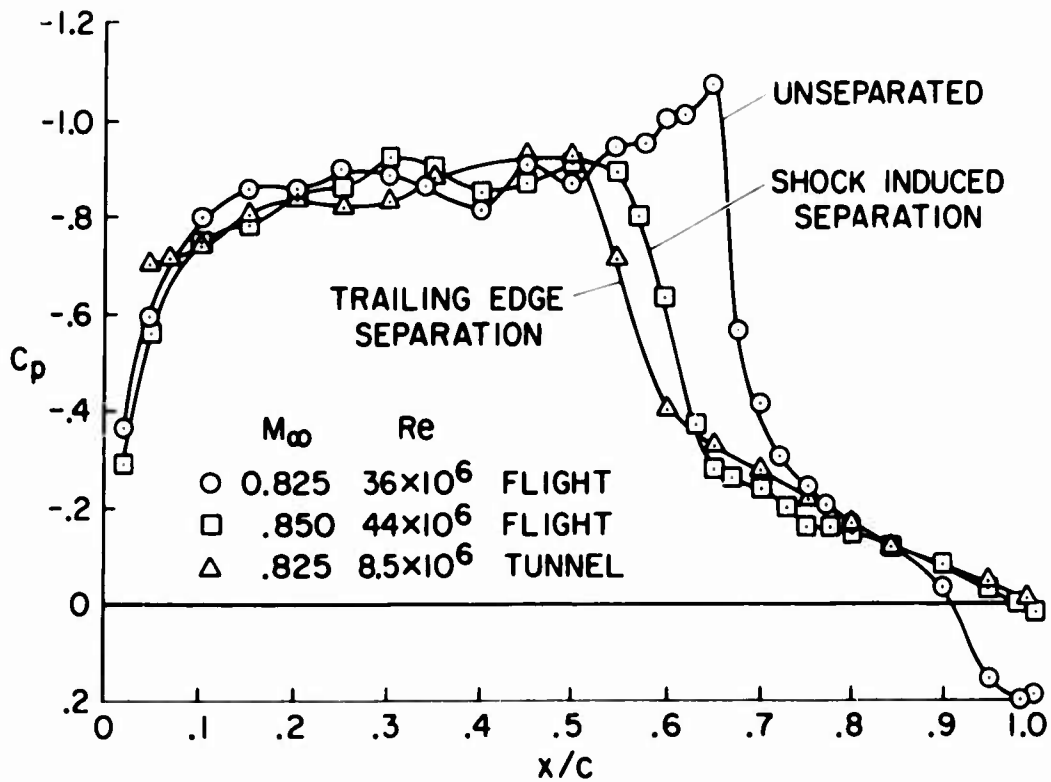


Fig. 5 Experimental surface pressure coefficients at 40% semi-span of the C-141

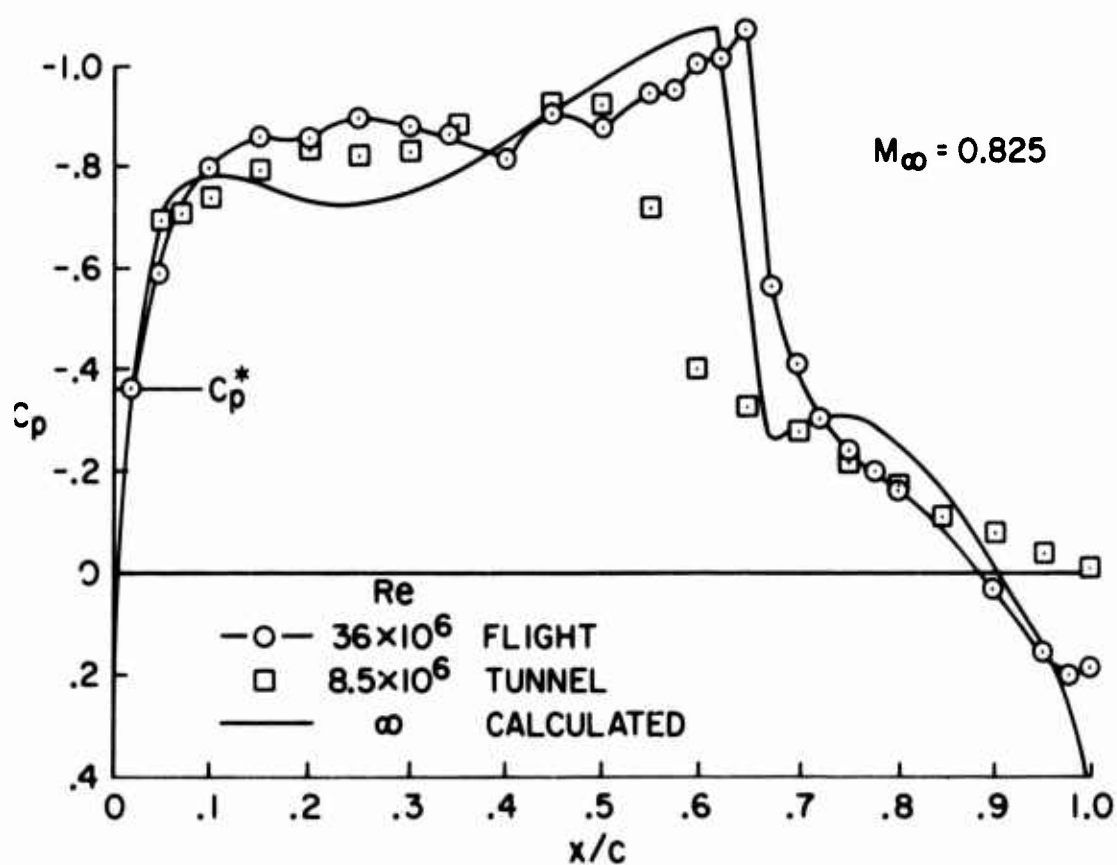


Fig. 6 Computed and experimental pressure coefficients at 40% semi-span of the C-141

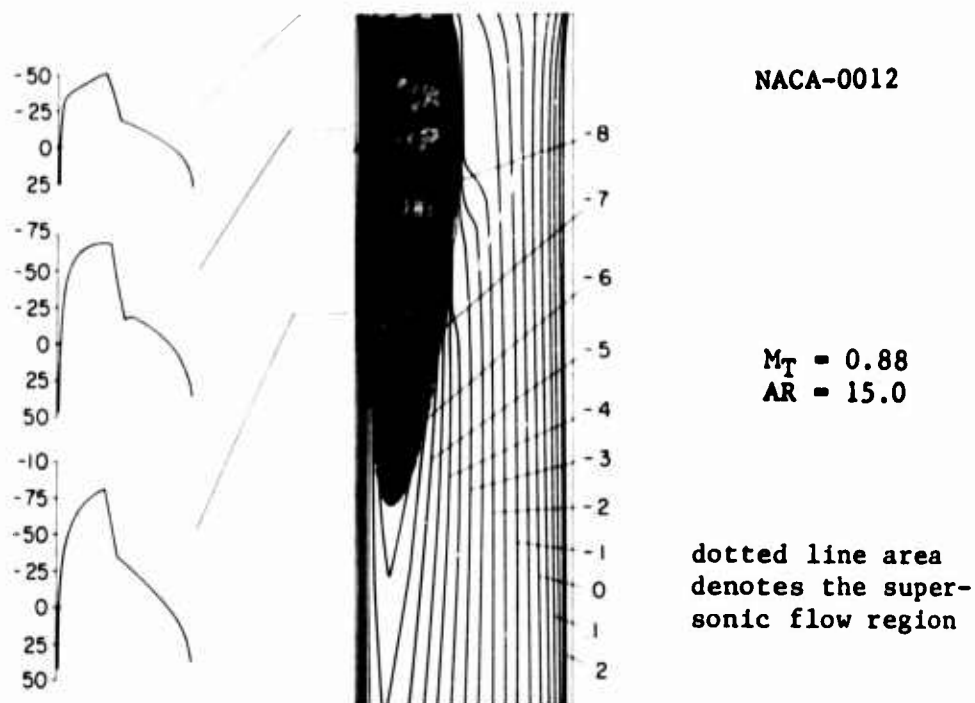


Fig. 7 Calculated pressure coefficients on a rectangular rotor tip

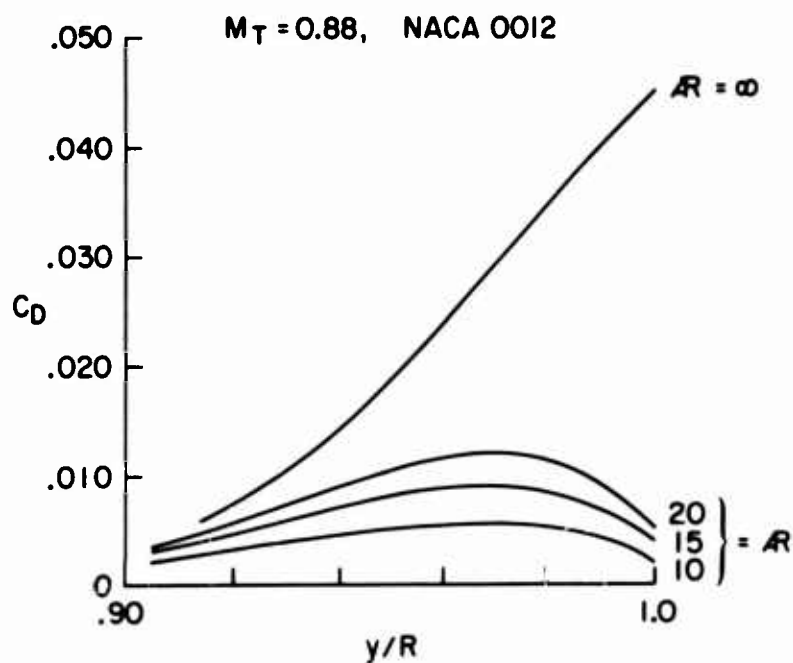
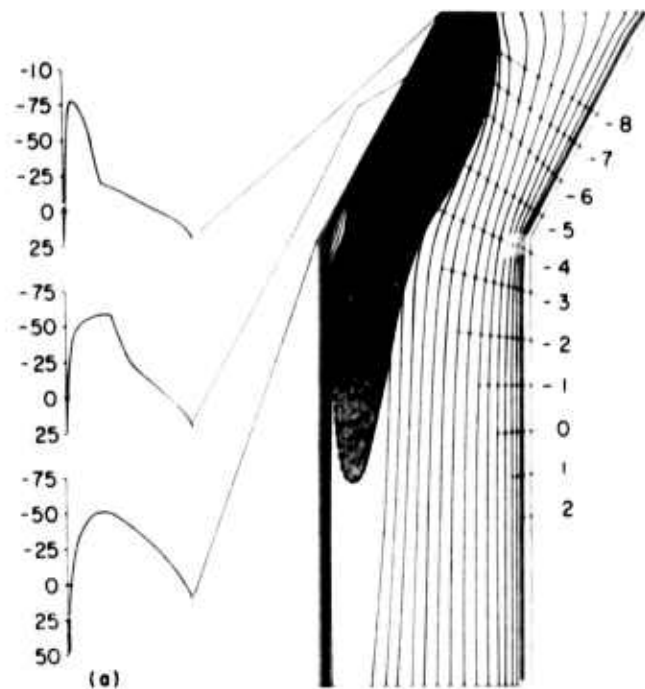
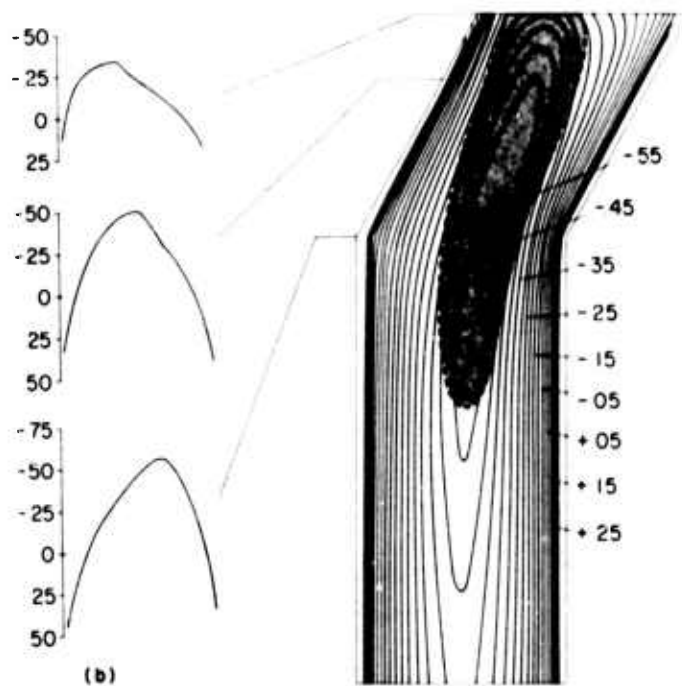


Fig. 8 Wave drag on several rectangular rotor tips



NACA-0012

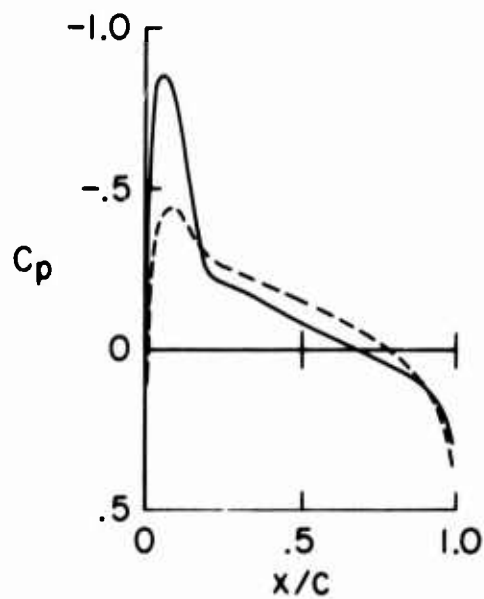


12% circular arc

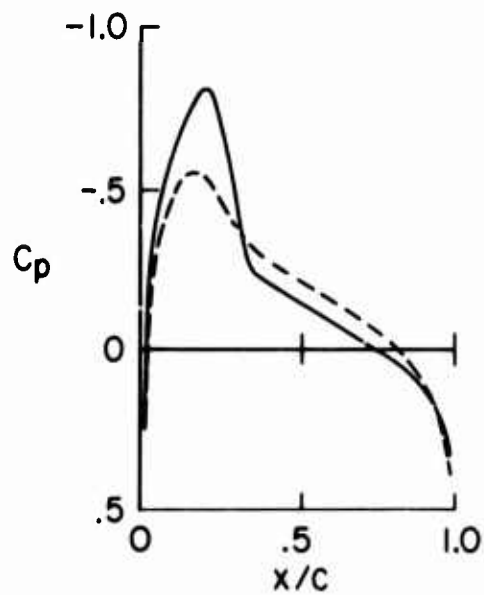
$M_T = 0.88$
 $AR = 15.0$

dotted line area
denotes the super-
sonic flow regions

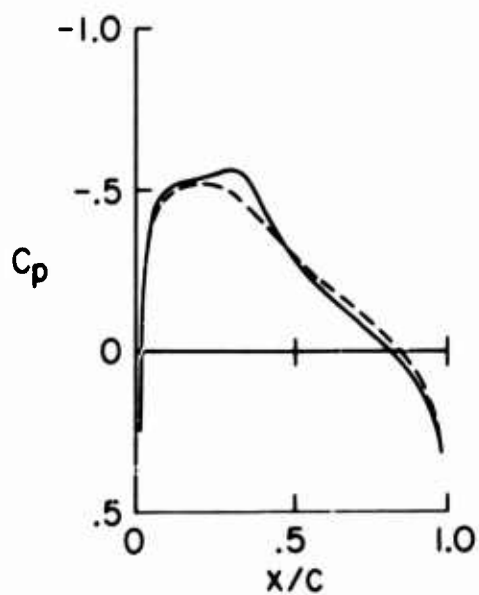
Fig. 9 Calculated pressure coefficients on swept rotor tips



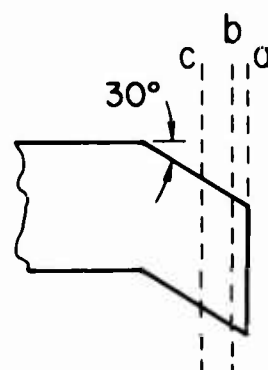
a) $y/R = 1.0$



b) $y/R = 0.99$



c) $y/R = 0.967$



— NACA 0012
 --- 12% "MIXED" PROFILE

Fig. 10 Calculated pressure coefficients on swept rotor tips

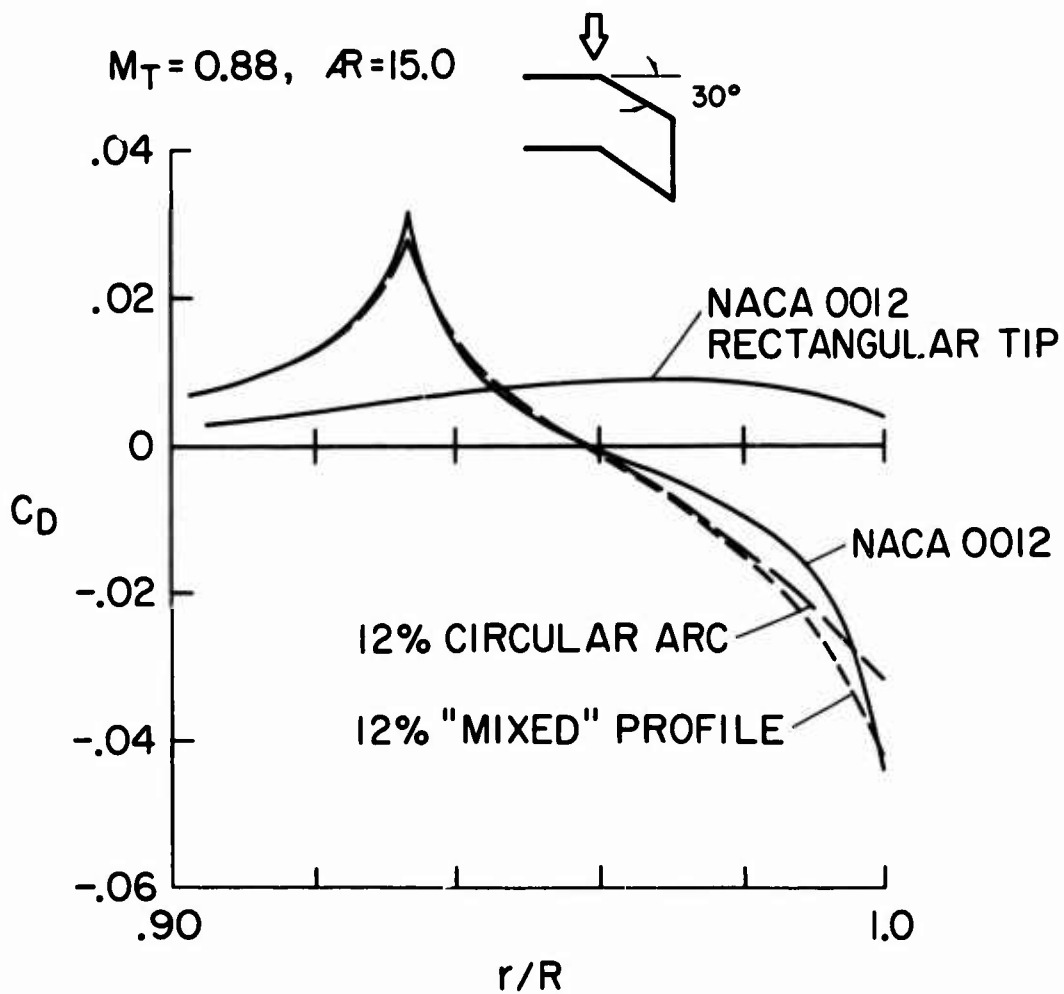


Fig. 11 Wave drags for rotor tips

AN ANALYTICAL EVALUATION OF AIRFOIL SECTIONS
FOR HELICOPTER ROTOR APPLICATION

GENE J. BINGHAM
LANGLEY DIRECTORATE
UNITED STATES ARMY AIR MOBILITY R & D LABORATORY
HAMPTON, VIRGINIA

INTRODUCTION

The NACA 0012 airfoil section was wisely selected for early helicopter rotor application and was used almost exclusively from 1939 to the mid 1960's. Alternate airfoils were hardly considered during this period because helicopter aerodynamic problems were secondary to the many mechanical and structural problems related to flight controls, power systems, and structural life. In about 1965, aerodynamic considerations resulted in the use of the NACA 23012 airfoil on one operational helicopter. From about 1965 to the present, a new effort has been made to define more effective airfoil sections for helicopter rotors and to identify a systematic approach to rotor airfoil design and selection (Refs. 1 through 6). As part of the continuing effort, an analytical investigation has been initiated to define the lift coefficient - Mach number characteristics of airfoils of potential interest for helicopter application. Then, this analysis was extended to determine the relative importance of the airfoil geometric parameters such as thickness, thickness distribution, leading-edge radius, camber, and camber distribution on lift-Mach number characteristics.

The analysis was based on the drag divergence Mach number prediction techniques of References 7 and 8. The airfoils considered were from the NACA four-digit, five-digit, and six-series families reported in Reference 9 and are generally consistent with the pitching-moment coefficient criterion suggested by Reference 6. Airfoil coordinates were derived by applying References 10, 11, and 12.

*Assigned to Langley Research Center, Airfoil Research Section.

BINGHAM

SYMBOLS

C	Chord of airfoil
C_d	Section drag coefficient
C_l	Section lift coefficient
$C_{m,ac}$	Section pitching moment about aerodynamic center
C_p	Static pressure coefficient
M_∞	Free-stream Mach number
M_d	Drag divergence Mach number, $\Delta C_d / \Delta M_\infty = 0.1$
R	Rotor radius
X	Distance measured along airfoil chord
X_c	Distance measured along airfoil chord from leading edge to crest
α	Angle of attack of airfoil (see Fig. 2)
ψ	Rotor blade azimuth angle

ANALYTICAL METHOD

BACKGROUND. On a single revolution, rotor-blade sections can experience lift coefficients from near zero to maximum lift and local Mach numbers from low subsonic to transonic values. Typical lift coefficient-Mach number envelopes are presented in Figure 1(a). The envelopes are shown for 0.6R, 0.8R, and 1.0R (the rotor tip) for a heavily loaded rotor flying at 136 knots (Ref. 5). At the rotor tip, the airfoil section advancing into the wind operates at Mach numbers to 0.9 at near zero lift coefficient and the retreating airfoil section operates up to a lift coefficient of 1.4 (near maximum lift) at Mach 0.45.

The drag divergence Mach number (M_d) for the airfoil section employed has been superimposed on Figure 1(a); the drag divergence Mach number is the free-stream Mach number at which the rate of increase of drag coefficient with Mach number equals 0.1. As noted on the figure, beyond about 0.7 to 0.8 radius, which includes over one-third of rotor area, the airfoil sections operate at Mach numbers

above drag divergence. Of course, the increase in drag has a prime influence on the power required to drive the rotor and thus the flight range and/or lift capability of the rotor. This point is even clearer when it is realized that the power absorbed by a rotor varies as Mach number cubed; the outer regions of the rotor are the major power absorbers. Therefore, an airfoil is desired which would have the drag divergence Mach number line located at as large a radius station (as far to the right) as possible.

The operating conditions for another helicopter configuration are presented in Figure 1(b). The calculated lift coefficient-Mach number characteristics only at 0.95 radius are shown for two vehicle weights at a flight velocity of 150 knots. In this case, the drag divergence Mach number will be exceeded only by the advancing blade at lift coefficients below 0.8. Although not shown on the figure, the drag divergence in this case extends inboard to about 0.8R. An analysis of this rotor indicates that each increase in drag divergence Mach number of about 0.01 would result in a 4-percent power savings for these operating conditions. For this analysis, all parameters except the drag divergence Mach numbers were unchanged. A larger power savings would be expected at higher flight velocities or with the configuration of Figure 1(a) because more of the rotor would operate beyond drag divergence. Because of these potential savings, one objective of the present analysis was to determine how to increase the drag divergence Mach number. More specifically, the intent was to determine the influence of the individual variations in airfoil geometric parameters on drag divergence Mach number, and then examine them in combination.

APPROACH. An analysis was performed to predict the drag divergence Mach number at lift coefficients from zero to near maximum lift. References 7 and 8 show that drag divergence can be predicted if the airfoil static pressure distribution and the location of the airfoil crest are known. The airfoil crest position and its variation with angle of attack is illustrated in Figure 2. Here the crest is identified as the chordwise station of which the airfoil surface is tangent to the free stream. Also, representative subsonic pressure distributions have been plotted for this airfoil and crest pressure coefficient has been identified.

Drag divergence results from changes in the airfoil surface pressure distribution caused by increases in Mach number. At free-stream Mach numbers below that for local supersonic flow, the pressure coefficients increase with Mach number according to the Prandtl-Glauert compressibility factor ($1/\sqrt{1 - M^2}$) and the drag coefficient is generally unchanged. As the Mach number is increased above that for

local sonic flow, supersonic flow develops in the low-pressure region of the airfoil, as illustrated in the sketch on Figure 2. The supersonic region contains alternate expansion and compression waves (Refs. 7 and 13) which change the surface pressure distribution. If the entire supersonic flow region is ahead of the crest, the change can cause the drag coefficient to increase, decrease, or remain constant. Increases, often called the creepy drag rise, result from net increases in pressures ahead of the crest; decreases observed by several investigators (see, for example, data of Ref. 14) result from net decreases in pressure ahead of the crest. As the Mach number is increased sufficiently to cause sonic flow to extend behind the airfoil crest, Reference 7 states that the expansion and compression waves cause the pressure coefficients ahead of the crest to become less negative and the pressure coefficients behind the crest become more negative which results in drag divergence.

Reference 15 suggests that a crest local static to free-stream total pressure ratio of 0.515 (Mach number of 1.02) is more appropriate to predictions of drag divergence than sonic Mach number at the crest. The Reference 15 criterion was applied to about 100 airfoils (Ref. 16) and it was concluded that the analytical results generally agreed with experimental drag rise Mach number results within ± 0.015 . In the present effort, sonic crest velocity has been used to predict drag rise Mach numbers because of the physical significance of the phenomena previously discussed and because the sonic crest approach is more conservative than the approach of Reference 15.

The lift coefficients also are influenced by the changes in pressure distribution with increasing Mach number. When supersonic flow develops on the airfoil, the reduced pressure coefficients in the supersonic region cause an increase in lift coefficient greater than would be predicted by the Prandtl-Glauert factor (Ref. 15). This increase continues until lift divergence which is evidenced by a reduction in lift coefficient with further increases in Mach numbers at a constant angle of attack. This reduction usually occurs at a Mach number slightly higher than the drag divergence value; that is, the reduced pressure region behind the crest that causes the drag rise can have a favorable influence on lift coefficient.

The pitching-moment coefficient variation with Mach number is not as clearly understood as the drag and lift coefficients. At Mach numbers below that for local supersonic flow, it might be expected that the pitching-moment coefficient would vary according to the Prandtl-Glauert factor; however, Reference 14 shows that this is not the case.

APPLICATION. The application of the analytical approach involves several steps: (1) low-speed pressure distributions, lift coefficients, pitching-moment coefficients, and crest location at selected angles of attack were determined by a compressible viscous flow mathematical model described in Reference 17; (2) the pressure coefficient at the crest was identified; (3) the Prandtl-Glauert compressibility factor was applied to the crest pressure coefficient to determine the free-stream Mach number at which sonic flow would exist at the airfoil crest (defined as drag divergence Mach number); and (4) the Prandtl-Glauert factor was applied to the computed lift coefficients to predict the lift coefficients at the drag divergence Mach numbers.

At present, the available aerodynamic computational tools do not permit predictions of the maximum lift coefficient because the computed results are not valid for cases with significant boundary-layer separation. In this investigation, results from the mathematical model with separation forward of 98 percent chord are excluded. In instances where the low subsonic Mach number maximum lift coefficient could be determined from experimental results (for example, from Ref. 9), the application of analytical results was carried to the known maximum lift coefficient.

As previously discussed, the pitching-moment coefficient does not follow that predicted by the Prandtl-Glauert factor. Therefore, because of an absence of an effective prediction criterion, the pitching-moment coefficients predicted by the mathematical model have been employed. Moreover, adequate predictions of drag by mathematical models are not currently available within the state of the arts and, therefore, are not used in this analysis. As suggested by Reference 6, an allowable pitching-moment coefficient of $|0.02|$ has been adopted in this analysis.

RESULTS AND DISCUSSION

The influence of the more important independent geometric variables (thickness, thickness distribution, leading-edge radius, camber, position of maximum camber) on drag divergence Mach numbers at various section lift coefficient have been predicted and are presented in Figures 3 through 8. Data are included for NACA four-digit, five-digit, and six-series airfoils to indicate the generality of the results.

THICKNESS VARIATIONS. The influence of variations in thickness-to-chord ratio are presented in Figure 3. The airfoils presented are symmetrical sections; the four-digit and five-digit

BINGHAM

airfoils have the same thickness distribution and, therefore, have the same aerodynamic characteristics. At zero lift coefficient, increasing the thickness-to-chord from 0.08 to 0.16 decreased the drag divergence Mach number about 0.08 for the four- and five-digit airfoils and about 0.10 for the six-series airfoils. The decrease was caused by an increase in the magnitude of the crest pressure coefficient with increases in thickness. As the maximum indicated lift coefficient was approached, the trends with increasing thickness reverse to result in a higher drag divergence Mach number for the thicker sections. In this case, the crest is farther aft for the thicker sections so that the pressure coefficient is of lower magnitude. The results of Reference 9 suggest higher maximum lift coefficients for the NACA four- and five-digit airfoils than for the NACA six-series airfoils of the same thickness ratio. The pitching-moment coefficient about the aerodynamic center of the symmetrical sections is 0.000.

The effects of changes in location of maximum thickness are presented in Figure 4. Results are presented for the four- and five-digit airfoils with maximum thickness ratio located from 30 to 50 percent chord as indicated by the last digit of the airfoil designation. The maximum thickness for the 63-012 and 65-012 sections is located at 35 and 40 percent chord, respectively. At zero lift coefficient, the drag divergence Mach number for the NACA four- and five-digit airfoils increased about 0.03 by moving the thickness location aft from 30 to 50 percent chord because of both the aft movement of the airfoil crest and the accompanying thinning of the leading-edge region which resulted in crest pressure coefficients of smaller magnitude. The increase was slightly greater at a lift coefficient of 0.4. At higher lift coefficients, the crest was more aft for the NACA 0012-63 section than for the other two airfoil sections and resulted in a crest static pressure of smaller magnitude, thus a higher drag divergence Mach number. The crest position of the NACA 65-012 section was farther aft than for the 63-012 section only at the lower lift coefficients. However, the pressure coefficient was of lower magnitude for the NACA 65-012 section at all lift coefficients analyzed (Fig. 4) to result in a consistently higher drag divergence Mach number than for the NACA 63-012 section. From this analysis, it appears that a maximum thickness location of about 40 percent chord is best from a drag divergence Mach number point of view. Again, the absence of camber results in a pitching-moment coefficient of 0.000.

LEADING-EDGE RADIUS. The influence of variations in leading-edge radius is indicated in Figure 5 only for the NACA four- and five-digit airfoils. The six-series airfoils were excluded because the leading-edge radius is not a design variable. The -33, -63, and -93 sections have leading-edge radius which are quarter

normal, normal, and three times normal, respectively (Ref. 10). At zero lift coefficient, the drag divergence Mach number increased about 0.02 when the leading-edge radius was increased from one-fourth to three times normal. This increase resulted from a decrease in magnitude of static pressure coefficient at the crest; the crest location at zero lift was the same for all three sections. At lift coefficients approaching the maximum, the crest moved rapidly forward as the leading-edge radius was increased and the pressure coefficient increased in magnitude to result in a decrease in drag divergence Mach number. An increase in leading-edge radius from one-fourth to normal had little influence on drag divergence for the NACA five-digit airfoils studied.

CAMBER. The effects of camber (Figs. 6, 7, and 8) were investigated only on the NACA five-digit and six-series airfoils because the type camber of the NACA four-digit airfoils produces excessive pitching moment for helicopter rotor application. The camber addition to the airfoils shown in Figure 6 is proportional to the design lift coefficients of 0.0, 0.3, and 0.6 for the NACA 0012-, 23012-, and 43012-63 airfoils and to the design lift coefficients of 0.0, 0.2, and 0.5 for the NACA 63-012, 63-212, and 63-512 airfoils, respectively (Ref. 9). The camber addition caused a decrease in drag divergence Mach number as great as 0.02 at zero lift coefficient. Since the angle of attack for zero lift coefficient decreased (more negative) as camber was increased, the airfoil crest moved aft. However, the magnitude of the crest pressure coefficient increased and resulted in a decrease in the drag divergence Mach number. For these airfoil sections at a given lift coefficient and the aft movement of the crest with increased camber continued for the full range of lift coefficients shown. And, the favorable influence of increased camber at the high lift coefficient resulted from crest pressure coefficients of smaller magnitude than that for the lower camber cases. Although the pitching-moment coefficient of the NACA 63-512 airfoil exceeded the $|0.02|$ criterion of Reference 6 by a factor of 2, the results have been included to discern the trends. To satisfy this criterion, the maximum design lift coefficient should be 0.25 instead of 0.5 because pitching-moment coefficient is proportional to design lift coefficient (Ref. 9). Also, observe that the NACA five-digit airfoils permit a higher design lift coefficient than the NACA six-series airfoils without exceeding the $|0.02|$ pitching-moment criterion.

The effect of varying the location of maximum camber is presented in Figure 7. The maximum camber is located at 5, 15, and 25 percent chord for the 21012-63, 23012-63, and 25012-63 airfoils, respectively, and at 32.5 and 50 percent chord for the 63-212 $a = 0.0$ and 63-212 $a = 1.0$ airfoils, respectively. These locations

represent the limits defined for the standard NACA five-digit and six-series airfoils (Ref. 9). At near zero lift coefficient, the results show that the position of maximum camber has little influence on drag divergence Mach number. The most significant influence is indicated for the five-digit sections at high lift coefficients. At these conditions, the most rearward position of maximum camber resulted in the most rearward crest location and thus the highest drag divergence Mach numbers. The aft movement of maximum camber also increased the pitching-moment coefficient, but it does meet the $|0.02|$ criterion of Reference 6 for the NACA five-digit airfoils presented. The pitching-moment coefficient of the NACA 63-212 $a = 1.0$ is excessive.

In addition to the conventional camber lines, the NACA five-digit airfoil series has been provided with a camber line which has trailing-edge reflex (Ref. 11). This reflex permits near zero pitching-moment coefficient at all lift coefficients up to the maximum lift coefficient. At a given lift coefficient, the reflex camber unloads the airfoil trailing-edge region and hence increases the magnitude of the forward upper-surface pressure coefficients compared to those for the conventional camber lines. This forward loading resulted in a corresponding decrease in drag divergence Mach number at all lift coefficients as shown in Figure 8; here the results for an airfoil with reflex camber (NACA 25112-63) are compared to those for an airfoil without reflex camber (NACA 25012-63). The decrease in drag divergence Mach number varied from 0.025 at zero lift to 0.01 at high lift coefficients.

COMBINED GEOMETRY. The influence of combining several of the favorable geometric parameters previously discussed is shown in Figure 9. The thickness distribution of the NACA four- and five-digit airfoils has been selected to take advantage of the low-speed maximum lift capability for the low pitching-moment designs discussed earlier. The influence of thickness ratio from 8 to 12 percent for these cambered sections is presented. Maximum thickness was located at 40 percent chord because, as previously discussed, this position was found to be superior to the 30-percent location at lift coefficients below 1.3 (Fig. 4). The standard leading-edge radius was selected because it was near optimum for the NACA five-digit airfoils (Fig. 5). Camber was added to increase the drag divergence at the higher lift coefficients; however, the camber was limited to satisfy the maximum pitching-moment criterion of $|0.02|$. The maximum camber was located at 35 percent chord because of the favorable trends shown with the aft maximum camber location (Fig. 7). The 35-percent chord maximum camber location is more aft than that provided by the standard NACA airfoils. The camber line employed was obtained from Reference 18 and was

apparently generated by extrapolating the NACA camber-line geometry. The results obtained with these sections are compared to the classic NACA 0012 section in Figure 9.

The 12-percent-thick combined geometry airfoil section has a higher drag divergence Mach number than the 0012 airfoil at all lift coefficients greater than zero. For these configurations, increases in drag divergence Mach number at zero lift can be obtained by decreasing airfoil thickness. As discussed earlier, and as shown in Figure 9, this will result in decreases in drag divergence Mach number at the higher lift coefficient shown. The final choice of an airfoil section for a given helicopter rotor should depend on an analysis of the airfoil section requirements.

CONCLUDING REMARKS

An easily applied analytical technique has been investigated for preliminary evaluation of airfoils for helicopter rotor application. The technique permits assessment of the influences of airfoil geometric variations on the drag divergence Mach number at lift coefficients up to near maximum lift. Analytical results presented in this paper indicated the compromises in drag divergence Mach number which result from changes in (1) thickness ratio, (2) location of maximum thickness, (3) leading-edge radius, (4) camber addition, and (5) location of maximum camber for NACA four- and five-digit airfoils and some six-series airfoils of potential interest for helicopters. An example of airfoil sections which combines several of the favorable geometric changes has been presented. Although the final selection of an airfoil, or combination of airfoils, for a helicopter rotor should be made on the basis of the specific performance requirements, the methods of analysis employed in this investigation can be rapidly and effectively used with high confidence during preliminary vehicle design and airfoil selection.

REFERENCES

1. Davenport, Franklyn J., and Front, John V.: Airfoil Sections for Helicopter Rotors - A Reconsideration. Proceedings of the 22nd Annual National Forum, American Helicopter Society, 1966.
2. Wortmann, F. X., and Drees, Jan M.: Design of Airfoils for Rotors. Volume, 1, Proceedings of Third CAL/AVLABS Symposium, 1969.
3. Benson, R. Geoffrey, Dadone, Leone V., Gormont, E., and Kahler, Gary R.: Influence of Airfoils on Stall Flutter Boundaries of

BINGHAM

Articulated Helicopter Rotors. 28th National Forum, American Helicopter Society, May 1972.

4. Pearcey, H. H., Wilby, P. G., Riley, M. J., and Brotherhood, P.: The Derivation and Verification of a New Rotor Profile on the Basis of Flow Phenomena; Airfoil Research and Flight Test. Royal Aircraft Establishment, Technical Memo Aero 1440, August 1972. (Presented at AGARD Conference on Aerodynamics of Rotary Wings, Marseilles, 13-15 September 1972.)

5. Reichert, G., and Wagner, S. N.: Some Aspects of the Design of Rotor-Airfoil Shapes. AGARD-CPP-111, Conference on Aerodynamics of Rotary Wings, Marseilles, 13-15 September 1972.

6. Kemp, Larry D.: An Analytical Study for the Design of Advanced Rotor Airfoils. NASA CR-112297, March 1973.

7. Nitzberg, Gerald E., and Crandall, Steward: A Study of Flow Changes Associated With Airfoil Section Drag Rise at Supercritical Speeds. NACA Technical Note 1813, 1949.

8. Nitzberg, Gerald E., Crandall, Steward M., and Polentz, Perry P.: A Preliminary Investigation of the Usefulness of Camber in Obtaining Favorable Airfoil Section Drag Characteristics at Supercritical Speeds. NACA RMA9G20, 1949.

9. Abbott, Ira H., von Doenhoff, Albert E., and Stivens, Louis S., Jr.: Summary of Airfoil. NACA Report 824, 1945.

10. Stach, John, and von Doenhoff, Albert E.: Tests of 16 Related Airfoils at High Speeds. NACA Report 492, 1934.

11. Jacobs, Eastman N., and Pinkerton, Robert M.: Tests in the Variable-Density Wind Tunnel of Related Airfoils Having the Maximum Camber Unusually Far Forward. NACA Report 537, 1935.

12. Ladson, Charles L., and Brooks, Cuyler W., Jr.: A Computer Program to Obtain Ordinates for NACA 6- and 6A-Series Airfoils. Proposed NASA TN, 1974.

13. Nieuwland, G. Y., and Spee, B. M.: Transonic Shock-Free Flow, Fact or Fiction. AGARD Transonic Aerodynamics Conference, Proceedings No. 35, September 1968.

14. Graham, Donald, Jr., Nitzberg, Gerald E., and Olson, Robert N.: A Systematic Investigation of Pressure Distribution of

BINGHAM

High Speeds Over Five Representative NACA Low-Drag and Conventional Airfoil Sections. NACA Report 832, 1945.

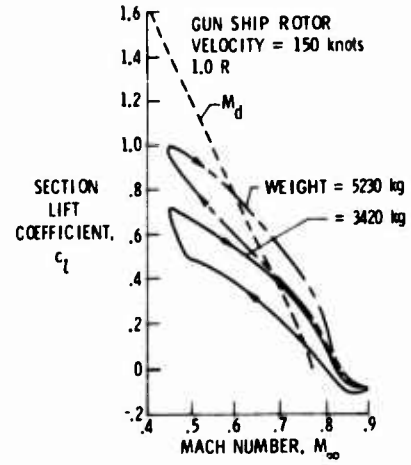
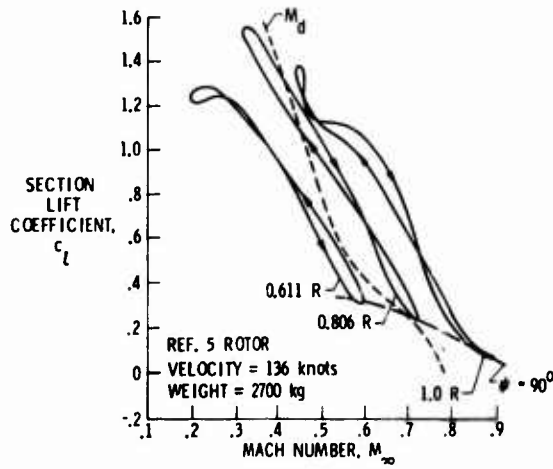
15. Sinnott, C. S.: Theoretical Prediction of the Transonic Characteristics of Airfoils. Journal of the Aerospace Sciences, pp. 275-283, March 1962.

16. Royal Aeronautical Society Transonic Aerodynamics Committee: A Method of Estimating Drag-Rise Mach Number for Two-Dimensional Airfoil Sections. Transonic Data Memorandum 6407, July 1964.

17. Stevens, W. A., Goradia, S. H., and Braden, J. A.: Mathematical Model for Two-Dimensional Multi-Component Airfoils in Viscous Flow. NASA CR-1843, July 1971.

18. Goethert, B.: High Speed Tests in the DVL High Speed Wind Tunnel on Airfoils of the Same Thickness Distribution With Different Cambers. Report and Translations N. 400, Ministry of Supply, 1947.

BINGHAM



(a) Heavily loaded rotor.

(b) Gun ship rotor.

Figure 1. Variation in lift coefficient during a revolution.

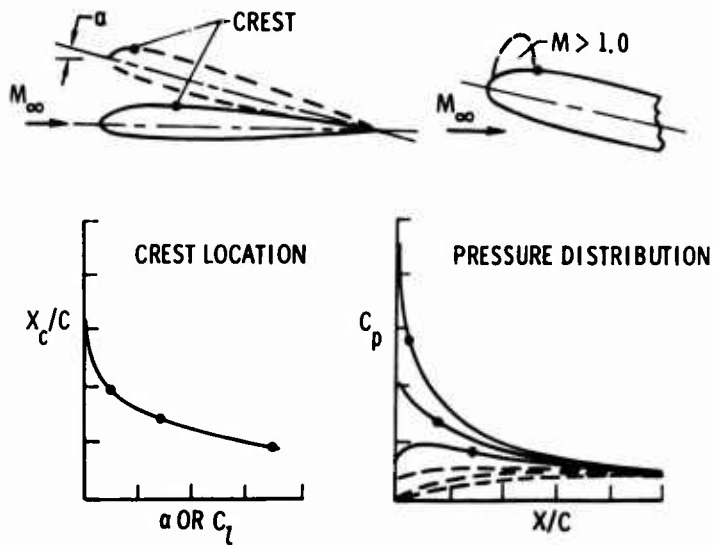


Figure 2. Representative crest location and pressure distribution.

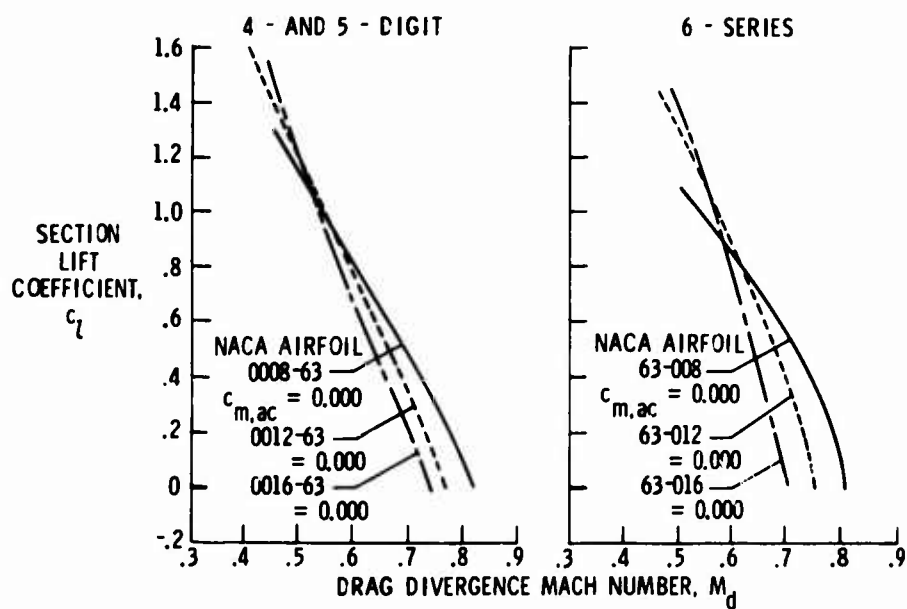


Figure 3. Influence at thickness ratio.

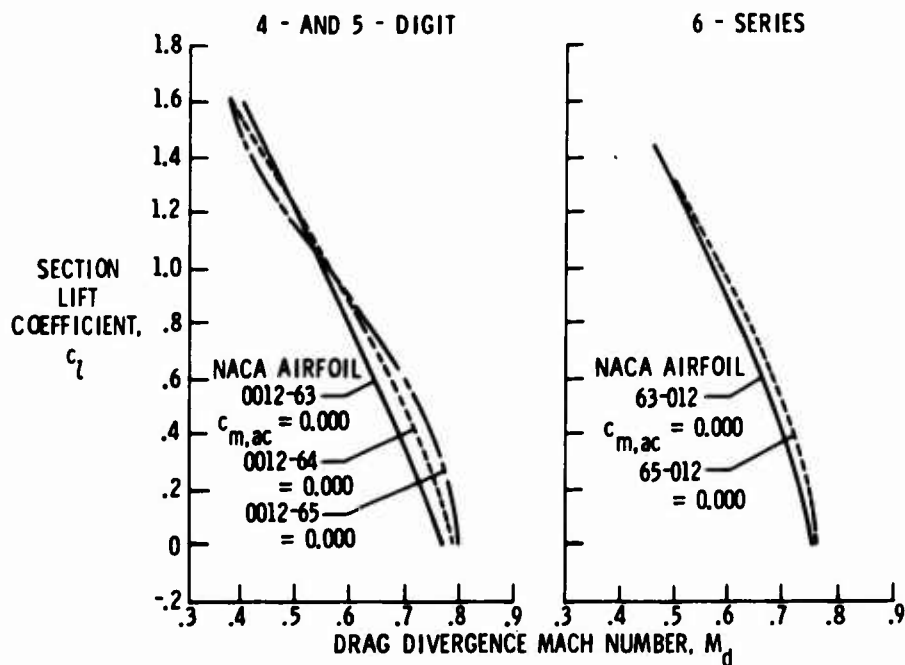


Figure 4. Influence of location of maximum thickness.

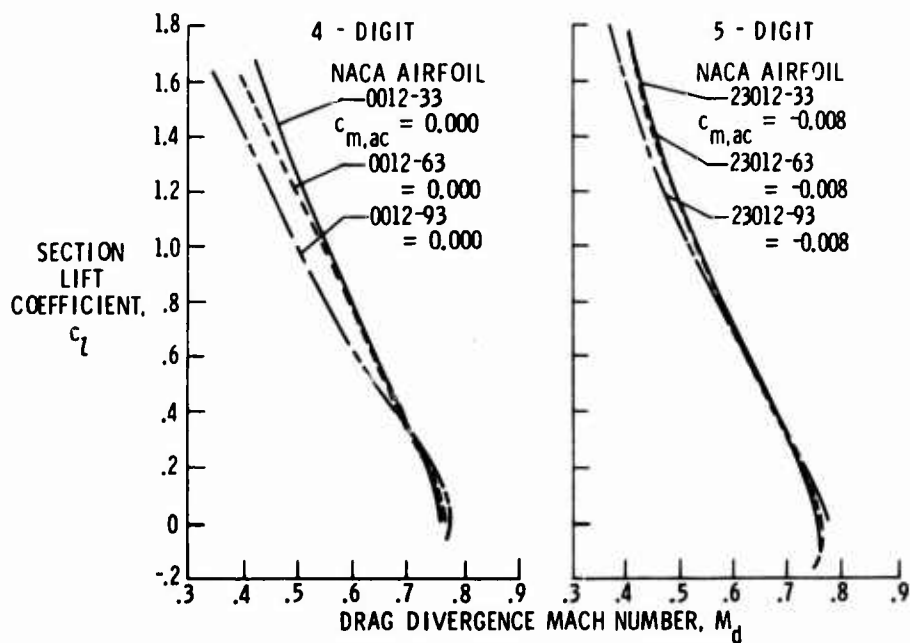


Figure 5. Influence of leading-edge radius.

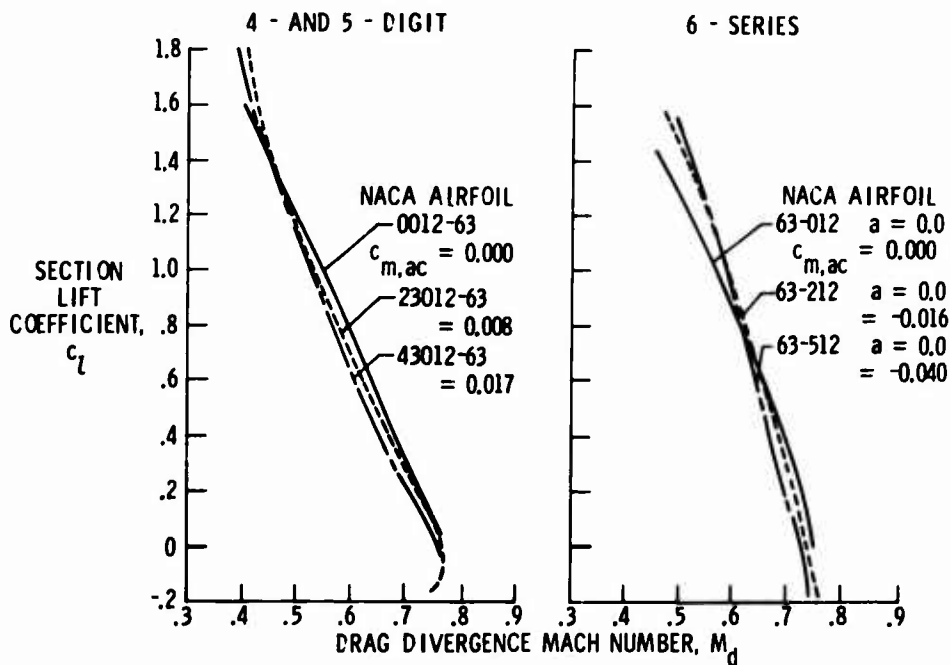


Figure 6. Influence of camber.

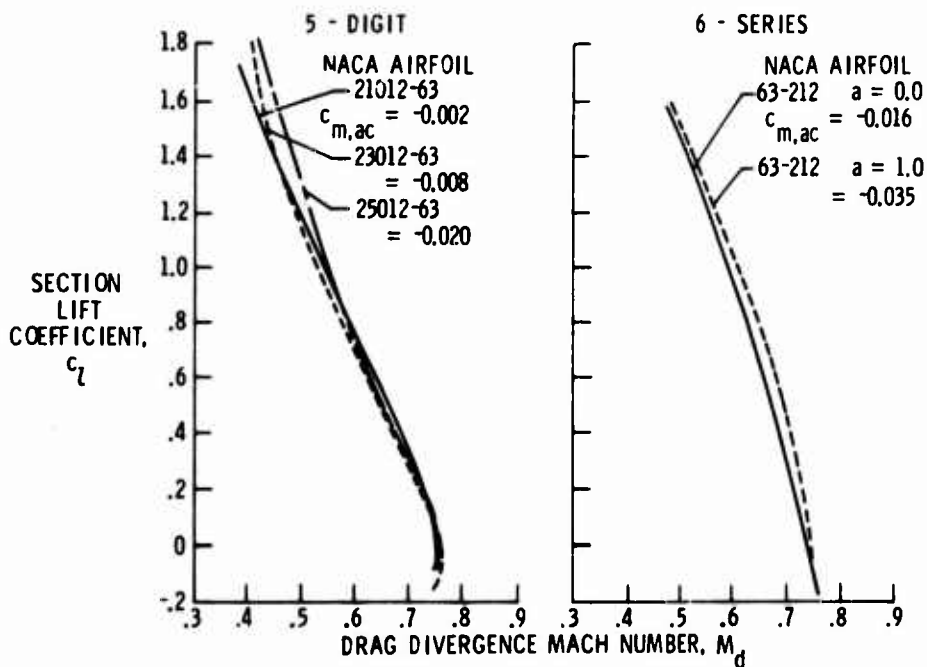


Figure 7. Influence of position of maximum camber.

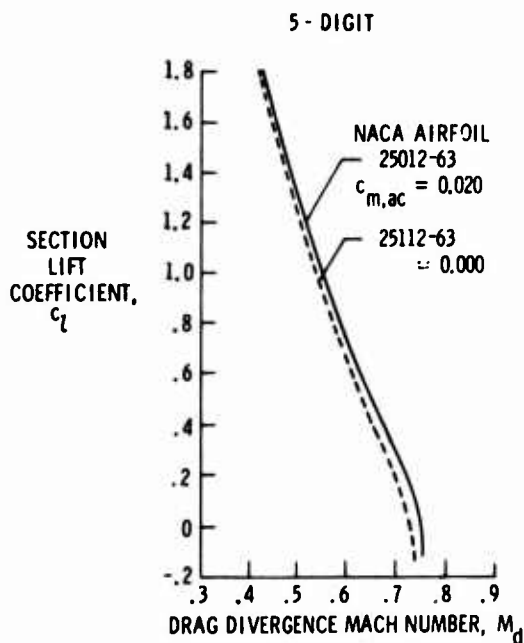


Figure 8. Influence of reflex camber.

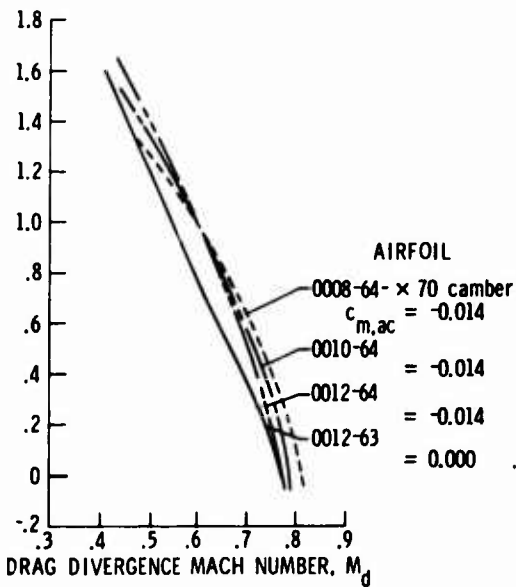


Figure 9. Influence of combination of geometric parameters.

HYPERVELOCITY IMPACT SHOCK-INDUCED DAMAGE TO STEEL ARMOR

JOHN W. BOND, JR., DR.
MOBILITY EQUIPMENT RESEARCH AND DEVELOPMENT CENTER
FORT BELVOIR, VIRGINIA

1. INTRODUCTION

An In-house Laboratory Independent Research (ILIR) program has been conducted at MERDC to design hypervelocity projectiles that maximize back-face spallation in steel armor. The projectiles are small (.5 to 5 gm each) so that several hundred of them can be located on one end of an explosive driver weighing 15 to 30 kg. The projectiles would be ejected in a near-vertical downward direction toward a target in a shotgun pattern at eject velocities of 8-10 km/sec and at a stand-off distance of 20 to 50 meters. This enables large lethal, or damage radii, of the order of tens of meters, to be considered. The primary target for the experimental program has been steel armor 2 to 5 cm thick. Damage would be by back-face spallation and fragmentation and resultant effects on soft interior components (of military vehicles) such as cabling, munitions, fuel, equipment, instruments, and personnel.

When the program started two years ago there was insufficient information on spallation physics to design the type of projectile desired. Accordingly, a program was started to obtain the needed experimental and analytical information. The specific goal was to design a projectile capable of shaping the pulse induced in the steel target in such a way that a large amount of steel was ejected from the back-face of the specimen in the form

BOND

of many small damaging fragments. During the course of the research, new phenomena were observed which are reported herein. In particular, the importance of the 130-kbar phase change in martensitic steel and its effect on spallation is discussed.

Over 100 successful impact experiments were performed in the Naval Research Laboratory (NRL) light gas gun facility. The experiments were designed and the results analyzed at MERDC. Metallurgical examination of the impacted specimens was done at the Stanford Research Institute (SRI), with some assistance from Sandia Laboratories Albuquerque (SLA). Mechanical material properties of the target materials were measured at MERDC.

In order to develop reliable theory and associated analytical tools, simple spherical projectiles, most weighing less than 1 gm, were used. Impact angles were generally normal, although a few non-normal impacts were made. A few impacts were made with both simple and composite non-spherical projectiles. Projectile materials included nylon, lexan, water, heavy oil, steel, MgLi, and several different ceramics

The reasoning behind these choices of projectile materials was rather complicated. It was based on the fact that the proposed warhead is weight limited rather than volume limited. This enables more projectiles of the same size to be carried, or else larger projectiles may be used. Size is important because the projectile diameter should be of the order of, or greater than the target thickness in order to optimize spall. There is some experimental evidence that much smaller (non-penetrating) projectiles are efficient spallators (3). (A spallator is a projectile that optimizes back-face spall. A penetrator optimizes penetration.) Finally, the vaporization or sublimation energy of the projectile material should be small compared to the kinetic energy of impact. The general idea is to convert most of the impact energy into induced shock. This means that inefficient processes such as penetration should be minimized.

Most of the targets were wrought, homogeneous steel plate stock 1.25 cm and 2.5 cm thick. The steel was a carbon

BOND

manganese-nickel-chromium-molybdenum alloy which was austenitized at 900°C, water quenched, and tempered at 540°C. The representative hardness was 360 Brinell (Bhn), and the approximate quasi-static tensile strength was 12 kbars. The dynamic spall strength corresponded to a critical tensile pressure of 38 kbars (1, 2, 3, 4).

After impact, the target specimens were dissected through the crater center, polished, etched, and examined metallurgically. Extensive 2-d code calculations of the impact spall phenomena were made by SLA for one specimen.

In Section 2, below, the experimental results are summarized. They are analyzed and discussed in Section 3. The 2-d computations performed by SLA are described in Section 4. The 130-kbar phase change and spall phenomena are discussed in the remaining sections.

2. EXPERIMENTAL RESULTS

Photographs of dissected impact craters are shown in Figure 1. The target material for Fig. 1a was 1.25 cm thick wrought steel. The projectile was a 0.52 gm nylon sphere with a diameter of 0.95 cm and an impact velocity of 5.18 km/sec. The crater is representative of craters formed by impact of lexan-encased H₂O and solid nylon spheres. The walls of such craters exhibit a series of evenly spaced ridges, or serrations, which are symmetric about a central axis of rotation. The macrocracks observed extend downwards from the bottom of the valleys in the crater floor. In Fig. 1b the target was 2.53 cm wrought steel and the projectile was a 0.7 gm steel sphere. In the case of steel-on-steel impacts the serrations in the craters are not observed and the macrocracks extend upwards. (This difference has not been explained.) In Figure 2 is shown a 10X photograph of the bottom of the center of a crater produced by a 6.03 km/sec impact of a 0.35 gm H₂O/lexan sphere on a 2.5 cm wrought steel target. The (apparent) shaded region extends completely around the crater. This region is readily visible on almost all impact specimens. It has extremely high hardness (500 Bhn) and displays a very fine untempered, martensitic microstructure. This observed metallurgical transformation has been shown to be due to the 130-kbar poly-

BOND

morphic phase transition in iron or martensitic steel (3, 4). The fine adiabatic shear lines are readily observable on this photograph. Also, note the large voids below the bottom of the crater.

When the stress wave, induced at impact, travels through the target and reflects in tension from the back-face, a spall layer, or gap, is produced as seen in Figure 1. Spall is not observed for these conditions at velocities below about 3 km/sec. At slightly greater velocities incipient spall is observed by the appearance of microcracks. The microcracks coalesce to form a complete spall layer at about 3.1 km/sec. As the impact velocity is increased the width and diameter of the spall layer also increases, and at some higher velocity the back-face plugs or fragments. Non-adiabatic shear occurs at the outer edge of the spall layer. For a 0.52 gm nylon sphere on 1.25 cm wrought steel, back-face plugging occurs at about 5.8 km/sec. For smaller (lighter) spheres, or for thicker targets, the plugging (or back-face fragmenting) velocity is higher.

In spite of the extensive experimentation described above (as well as all other experimentation performed in the US) it is still not possible to design an effective spallator. However, a few preliminary guidelines have been outlined. Briefly, it is clear that pulse shaping is the key to the production of damaging back-face spall. This is exemplified in Figure 3 which shows the reflection and resultant spall of a triangular stress pulse. The incident pulse reflects in tension at the free surface x_0 . Upon exceeding the spall threshold, a spall layer is formed at x_1 . The distance $x_1 - x_0$ must be sufficiently small so that the back-face can be spalled and fragmented, but sufficiently large so that a damaging amount of back-face material is ejected.

After the spall layer is formed at x_1 , the subsequent stress pulse reflects at x_1 rather than at x_0 . This can result in the formation of a second spall layer at x_2 . Obviously this sequence of events can be repeated to form multiple spall layers. In order to prove this hypothesis several different projectile configurations were tested. One of these configurations was a hollow plastic sphere. The reasoning here was that upon impact

BOND

more mass would be delivered at the impact periphery and "double-impact" conditions for thin flyers would exist at the impact center. This heuristic prediction was eminently borne out in several experiments in which the projectiles were hollow spheres. An example is shown in Figure 4. In this case the projectile was a hollow lexan sphere weighing 0.56 gm with a diameter of 1.08 cm and an impact velocity of 5.59 km/sec. Note the double spall layer and the back-face fragmentation.

A few other impact experiments were performed in which the projectile was computer designed to optimize back-face spall. These experiments were particularly successful in showing that back-face spall can be optimized.

3. DISCUSSION OF EXPERIMENTAL RESULTS

In general a well-conceived experimental program produces results that have not been predicted theoretically. This program is no exception. The experimental "surprises" are too numerous to cover in this report; but a few of the more important results can be discussed. These are listed below:

a. Serrations are formed at the crater floor for nylon-and water-steel impacts. These have not been explained; in particular, the reason for the serration formation on nylon-steel impact and the lack of serration formation on steel- steel impact has not been explained.

b. Preliminary or tentative explanation of the macrocracks observed for nylon-steel impacts has been given. However, this explanation fails to show why the macrocracks extend upwards for steel-steel impacts as compared to downwards for nylon-on-steel impacts.

c. Adiabatic shear lines are observed below the crater floor in almost all cases. However, their role in the damage process has not been explained.

e. The importance of the observed 130-kbar phase change has been explained for one case. This is discussed in the next

BOND

section. Much more work is needed on this phenomenon.

f. Spallation. The 2-d computer calculations performed by SLA have shown excellent agreement with one impact experiment. This applies to the configuration and location of the spall layer for a relatively simple case. If this work could be extended to more important and relevant cases it would be of extreme significance.

g. General Summary

(1) Metallographic/Fractographic Features

As indicated previously considerable circumstantial evidence leads to the conclusion that the dark-etching hemispherical volume of material under the impact site, such as shown in Fig. 2, has undergone the pressure-induced phase transformation:

$$\alpha \rightleftharpoons \epsilon$$

Material within this zone is of significantly finer grain structure and higher hardness than material elsewhere. Back surface laser interferometer records obtained by SLA in plate slap experiments exhibit clearly the disturbance attributable to a wave reflected from a denser phase (4).

Profuse shear bonding was observed in the subcrater region and is felt to play an important role in the cratering process. The long cracks extending into the armor steel from the crater walls tend to lie on shear bands and tend to join up to isolate pieces of material that form part of the ejecta.

Shear bands that are not associated with the crater walls are also prominent in the transformed region. A number of equiaxed microfractures having a ductile appearance are generally present in this region, linked together by adiabatic shear bands, which are easily observable because of their distinct etching behavior. Again, cracks are observed to have formed in the bands, and suggest the mechanism of material removal in this region. No evidence of adiabatic shear was found in the back surface regions.

BOND

Fracture damage in the back surface region proceeds by the nucleation, growth, and coalescence of microfractures, followed by a widening of the resulting macrocrack and subsequent scabbing of the back surface by a shearing process that does not appear to be adiabatic. Observations of back surface damage at all stages of development were possible because of the wide variations in impact conditions. However, detailed stress histories were not obtained.

(2) Correlations with Stress History

An attempt was made to correlate metallographic and fractographic features observed in the specimens with stress histories. This is described in the next section of this report. However, many of the essential features of the stress history may be similar in other experiments. Thus the correlations may be common to hypervelocity impact in general, although many more experiments, metallurgical analysis, and theoretical extensions are needed before this statement can be properly delineated. The observed features for the single case described above were predicted (post-test) with outstanding accuracy by the SLA.

4. 2-D COMPUTATIONS

It was fortuitous that the MERDC impact spallation experiments began about one year before SLA scheduled for usage their new two-dimensional hydrodynamic codes. Accordingly, in May 1973, SLA began to code the impact of a nylon sphere on to 1.25 cm steel armor. The experimental results are shown in Figure 1. The impact velocity was 5.18 km/sec and the projectile mass 0.52 gm. The results of the SLA effort are described below (4).

Two-dimensional code calculations of the above test were performed at SLA using both the CSQ Eulerian Code and the TOODY Lagrangian code with rezoning. The initial impact pressure was about 400 kbar which causes iron and martensitic steel near the impact point to undergo the $\alpha \rightarrow \gamma$ polymorphic phase change. The nylon sphere greatly distorts and goes into a liquid-partially vapor state. The steel plate suffers a large distortion near the impact point and requires both an accurate elastic-plastic model and a good material failure model elsewhere in the plate. Also, a high degree of resolution is

BOND

required for predicting complete or internal spallation because of the attenuation of the initial compressive pulse, its reflection, and the interaction of the compressive and the reflected release wave must all be calculated accurately.

The 130-kbar phase change, the correct spall strength, and an adequate failure model have all been determined to be very important for predictive and interpretive calculations. A detailed examination of the stress histories for the cases with and without the phase change has shown why different failures should be expected in these two cases (4). When the phase change is included in the calculations, the loading portion of the stress wave consists of two parts as the wave separates at 130 kbar. More important is the difference observed as unloading occurs; a rarefaction shock is present when the phase change is included. These differences in the stress wave structure not only cause a cylindrical-conical failure to occur directly below the crater, but also result in a propagating pulse that is nearly square for the case with a phase change and nearly triangular without. The square pulse, upon reflection from a free surface, transfers essentially all the momentum to the spall layer, whereas the triangular pulse is not nearly as effective in momentum transfer. This clearly shows the importance of the stress pulse shape in producing effective spall.

The measured spall stress of 38 kbar (2, 4) for the MS 12560 steel was used in the above calculations, and the results showed excellent agreement with the experiment. This was especially true for the Lagrangian results where the crater diameter, crater depth, spall layer thickness, spall length, spall bulge, cylindrical-conical failure, and the area that undergoes a phase change all show nearly one-to-one correspondence with experiment. The calculation did not, however, reproduce the serrations in the crater floor, nor the macrocracks and adiabatic shear observed below the crater floor.

When the phase change is included in both the CSQ and the TOODY calculations, the only significant difference is the spall length (or diameter). This difference is a direct result of the material failure treatment at the spall plane where a material coordinate treatment allows a higher accuracy. The very good numerical and experimental agreement demonstrates the capability of the codes to solve a wide class of difficult and important problems although

BOND

much more correlation between calculations and experiments is needed.

5. THE 130-KBAR PHASE CHANGE IN MARTENSITIC STEEL

In recent years a number of measurements have been made of the high pressure properties of iron in the region of the $\delta \rightarrow \epsilon$ phase transition. These indicate that at room temperature the transition occurs in the region of 130 kbars and that it is a baric, initiating at about 130 kbars and going to completion at pressures greater than 170 kbars. The observed specific volume change associated with the transition is $\Delta v \approx 0.0066 \text{ cm}^3/\text{gm}$ and the compression ratio at the onset of the transition is $\chi \approx 0.943$. Here χ is the specific volume, v , divided by the specific volume, v_0 , of δ at 0°K and zero pressure (6).

The phase diagram of Fe is shown in Figure 5. According to the SLA calculations the shock temperature in the Fe does not go sufficiently high to drive the steel into the γ phase. However, the shock pressure does get sufficiently high to drive the steel into the phase from which it reverts, on cooling, to the α phase, with a consequent change in grain structure and other mechanical properties. Direct active measurements to show this change have not been made but a combination of theory, computations, and metallurgical analysis indicates that the above history is correct.

6. COMPLETE AND BACKFACE SPALL

Microcracks and voids begin to form when the tensile stress reaches the "spall threshold" or the incipient spall strength. For practical purposes a certain microcrack size or density is assigned to this value.

As the tensile stress exceeds the spall threshold the microcracks begin to coalesce until a complete spall layer is formed. For the impact experiments with spherical projectiles described in this paper, the spall layer is generally parallel to the backface of the targets with its center line at a distance (d) from the backface. For a given type of projectile d is constant over the velocity range from 3 to 7 km/sec. The spall layer has a thickness δ (perpendicular to the backface), which increases rapidly with velocity above the incipient threshold. In addition to the spall layer thickness there is a backface bulge with height (h), which is equal to δ . As long as the spall layer is formed near the backface h is a useful experimental (or empirical) parameter since

BOND

it can be measured with good accuracy, whereas measurements of δ are sometimes ambiguous. The spall layer has a diameter D (parallel to the backface) that increases slightly with impact velocity. It grows to be somewhat greater than the crater diameter (parallel to and at the target surface). At some high impact velocity the backface "plugs" or fragments, i.e., material of thickness d and (approximate) diameter D is ejected from the backface either as a single plug, or as a number of small fragments. The impact velocity to produce incipient spall is V_0 and for backface spall it is V_∞ . As will be seen V_∞ appears to be about twice V_0 for the experiments described.

Backface bulge height h is plotted on linear paper as a function of impact velocity for three different cases in Figure 7. It is seen that the experimental points lie reasonably well on smooth curves. The points have been plotted for the .52-gm nylon sphere on the 1.25-cm steel target on semi-log paper in Figure 8. They fall well on a straight line which suggests that the fractional increase in h (or δ) increases with incremental velocity, i.e.,

$$\Delta h = k \delta \Delta V$$

where k is a constant. Upon integration this becomes

$$h = h_0 \exp k(V - V_0)$$

For the purposes here h_0 can be assumed to correspond to the bulge height at incipient spall, with V_0 the corresponding impact velocity.

REFERENCES

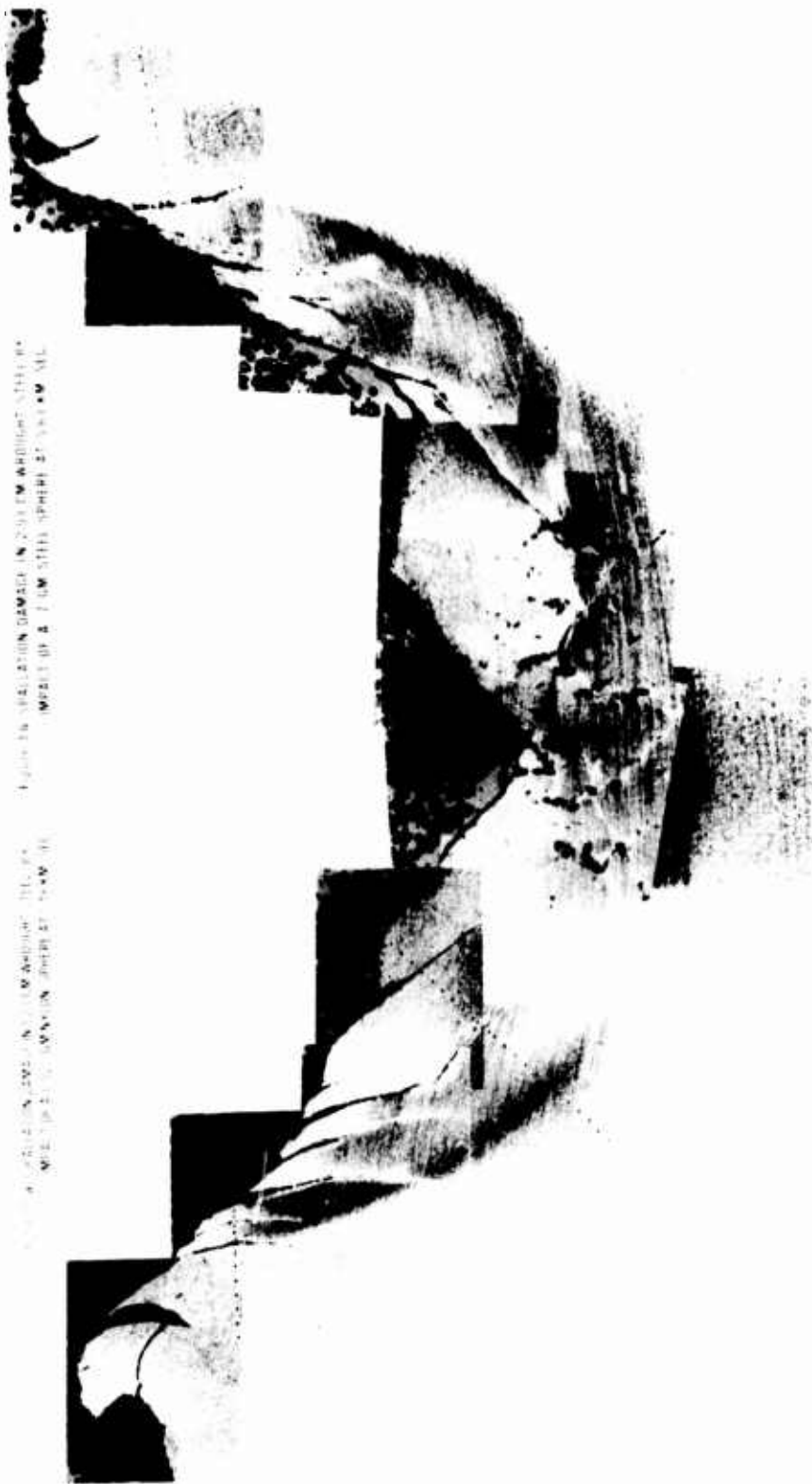
1. Two-Dimensional Spallation Induced by Hypervelocity Impact in Wrought Steel Plate, J.W. Bond, and G.W. Ullrich, MERDC report 2067, July 1973 (U)
2. Personal Communication, A. Stevens, SLA, 1973 (U)
3. Support of Armor Fragmentation Studies, D.A. Shockey, SRI Project PYU 2151, 16 Oct 1973 (U)
4. Hypervelocity Impact and Associated Phenomena, D.A. Shockey, SRI Project PYU 2151, 19 Dec 1973 (C)
5. Metallurgical Effects at High Strain Rates, D.J. Pastine, Plenum Press 1973.
6. Equation of State of the Alpha and Epsilon Phases of Iron, D.J. Andrews, Washington State University, WSU SDL 70-05, Nov 1970.
7. Dynamic Mechanical Behavior of Iron, D.R. Curran in Shock Waves and the Mechanical Properties of Solids, ed J.J. Burke and V. Weiss, Syracuse Union Press, 1971.

BOND



Figure 10. Radiation damage in 200 MW and 100 MW
MREI in a 1.5 M steel sphere at 100 MW.

Figure 11. Radiation damage in 200 MW and 100 MW
MREI in a 1.5 M steel sphere at 100 MW.



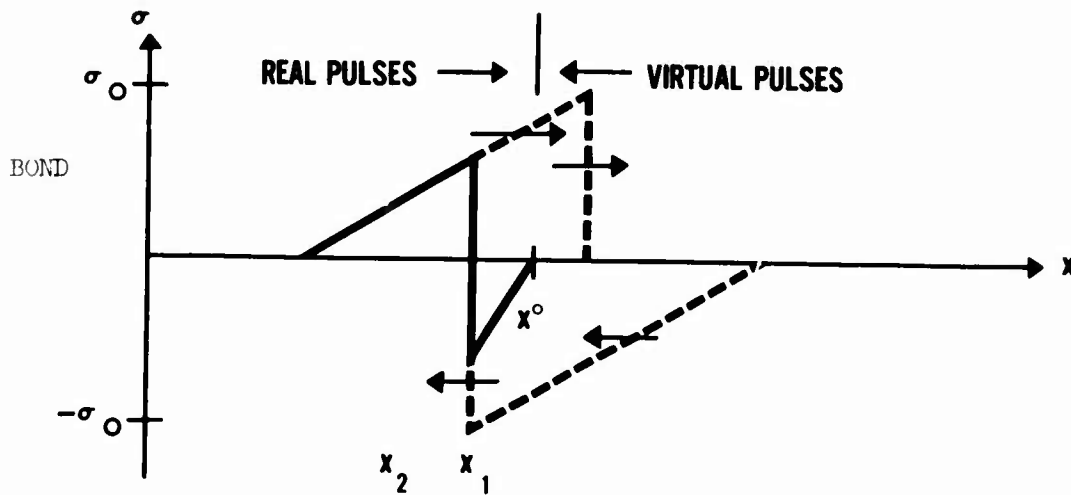


Figure 3. SPALL LAYER FORMED BY A TRIANGULAR PULSE



**Figure 4. SPALLATION DAMAGE IN 1.25-CM WROUGHT STEEL
BY HOLLOW LEXAN SPHERE. MASS=.565 GM,
DIAM=1.08 GM, IMPACT VELOCITY=5.6 KM/SEC**

BOND

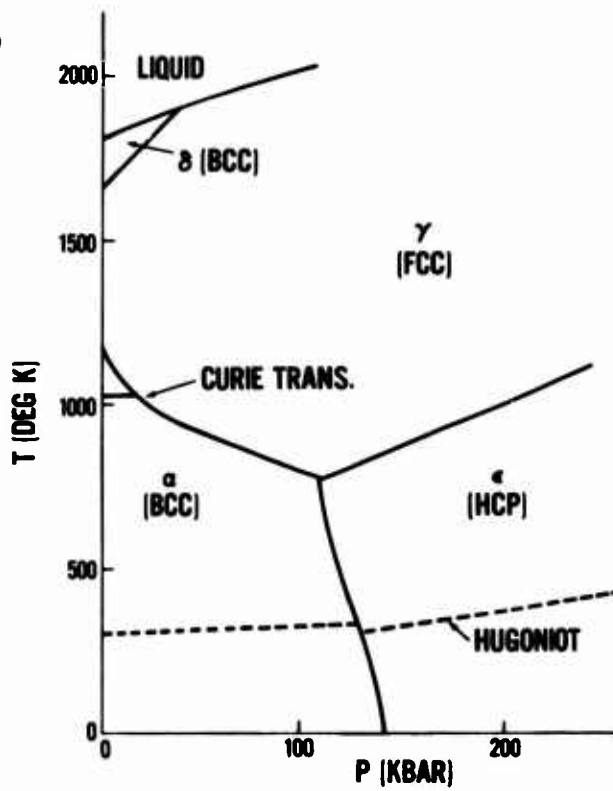


Figure 5. PHASE DIAGRAM OF IRON

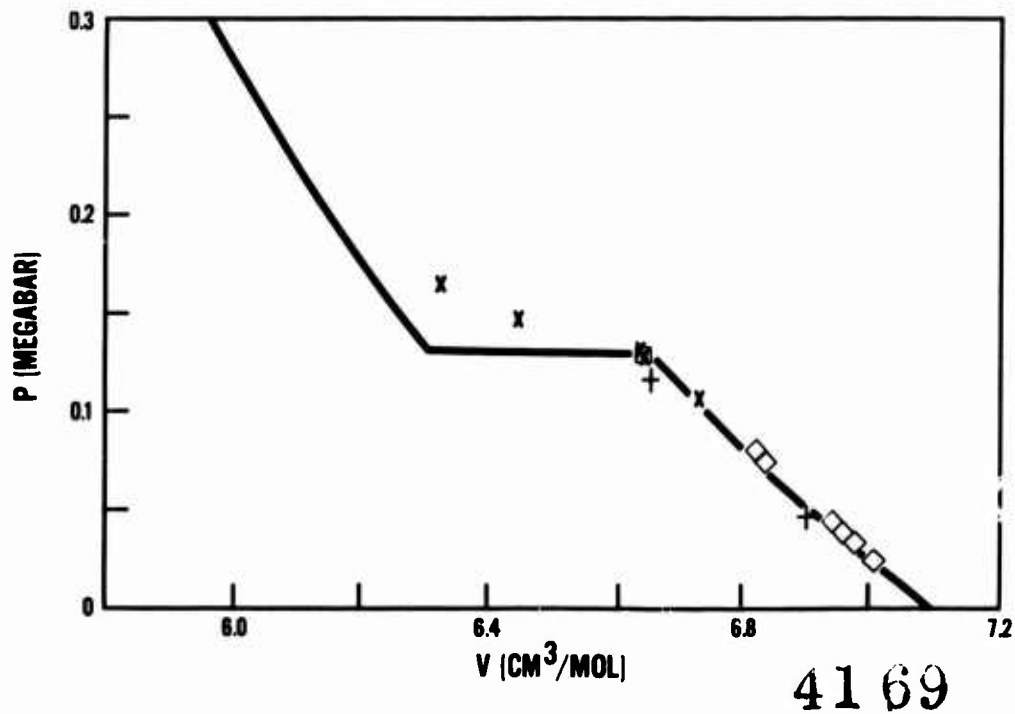


Figure 6. HUGONIOT STATES OF IRON

BOND

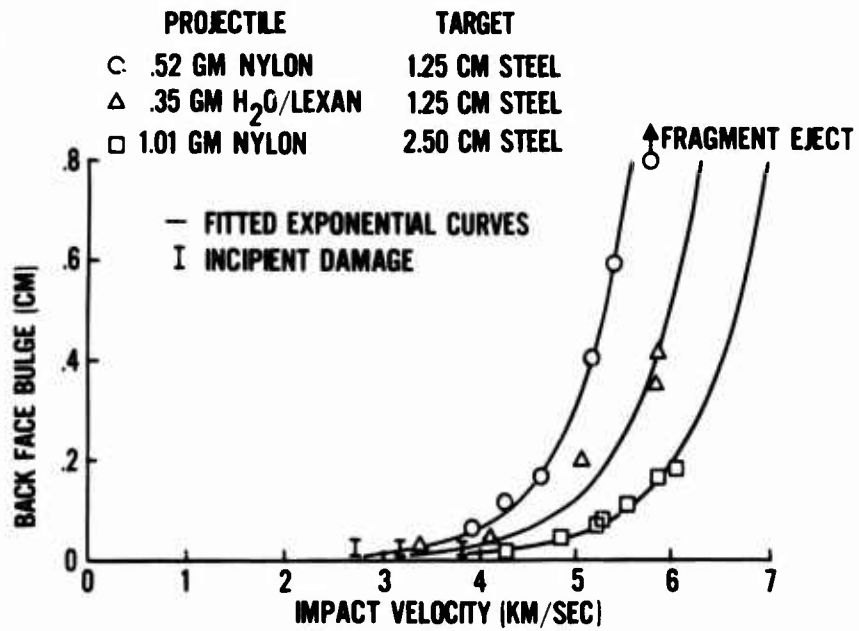


Figure 7. SPALL DAMAGE IN HOMOGENEOUS WROUGHT STEEL PLATE

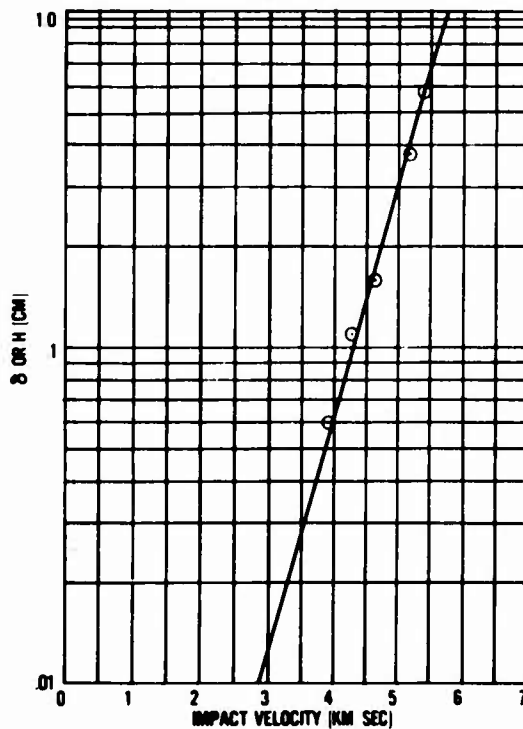


Figure 8. BACKFACE BULGE (OR SPALL LAYER THICKNESS) VS IMPACT VELOCITY

LARGE POLARIZATION DEPENDENT VOLTAGES IN FERROELECTRIC CERAMICS

PHILIP S. BRODY & FRANK CROWNE
HARRY DIAMOND LABORATORIES
WASHINGTON, D.C.

INTRODUCTION

This paper describes a new high-voltage photovoltaic effect (Ref. 1) appearing in ferroelectric ceramics. A thin ceramic wafer as shown (Fig. 1) illuminated uniformly produces a steady voltage across electrodes on the wafer edge.

The open-circuit voltages are large and proportional to the length, l , of the wafer. The exact voltage per centimeter length depends on the composition of the material and the grain size of the ceramic. The largest we have measured is 1500 volts per centimeter in a certain PLZT ceramic composition. Other open circuit voltages range from this value down to 350 V/cm for BaTiO_3 ceramic. Except for these very high voltages, the effect is similar to that in a p-n junction or a Schottky barrier photocell with a high output resistance operating in the unbiased (self generating) mode. If a resistive load is connected between the electrodes, a small steady current flows as long as the wafer is illuminated. Though the current is small, the voltage (across a high resistance) is high. Photovoltaic conversion efficiencies as high as 0.02% are obtained for a certain PZT composition with band gap (373-nm) illumination.

The short-circuit photocurrent is linear with intensity and proportional to the width. The open-circuit photovoltage is, at low levels, linearly dependent on intensity, saturating at moderate levels. The voltage and short circuit current are proportional to the remanent polarization of the ceramic wafer, vanishing when the polarization vanishes.

A much lower photovoltaic effect was seen previously in unpolarized ferroelectric ceramic (Ref. 2) and also in single crystals of the ferroelectric barium titanate (Ref. 3, 4). High-voltage photovoltaic effects in materials other than ferroelectrics have also been observed. These are seen in films of evaporated semi-conducting materials (Ref. 5). The effect has been attributed to an arrangement of p-n junctions in which small photovoltages add (Ref. 6). Voltages up to 100 V/cm are seen. A high-voltage photovoltaic effect has also been seen in insulating single crystals of zinc sulfide by Merz et.al. (Ref. 7). This is thought also to be a multi-junction effect. Stacking faults separating hexagonal and cubic modifications of the crystal have been suggested as the source of small voltages which then add in series (Ref. 6).

In the next section we will describe experimental results covering various materials, all ferroelectrics of the perovskite oxide type.

RESULTS AND DISCUSSION

Measurements were made using a high-pressure mercury arc with filters to produce illumination at various wavelengths and intensities. For temperature control, the samples were immersed in silicone oil, the temperature of which was controlled, or in a specially built thermal chamber. High open-circuit voltage measurements were made using a self-balancing potentiometric system drawing negligible current. The input impedance of the measuring circuit was thus orders of magnitude higher than that of the source impedance. Measurements of lower voltages were made using a vibrating capacitor electrometer. Short-circuit measurements were made with electrometers operating as ammeters. Thermopiles determined intensities.

The wafer as shown in Fig. 1 can be represented by an equivalent circuit consisting of a voltage in series with a high resistance. This is demonstrated by the results illustrated in Fig. 2, where an illuminated wafer was placed in series with a low source resistance voltage source applying voltage V_r across the electrodes (Fig. 3). Open-circuit voltage V_o is 630 volts, the intercept of the roughly straight line with the abscissa; the short circuit current, $i_{sc} = -.31 \times 10^{-8}$ A is the intercept with the ordinate. The photoresistance R equal to V/I is $2.2 \times 10^{11} \Omega$. In general, the output of a sample is well defined by designating the illumination (intensity and spectral distribution), open circuit voltage V_o , and short-circuit current i_{sc} . These quantities are shown for various materials in Table I. We define photovoltaic

efficiency as the power dissipated in a short circuited wafer divided by the incident power. All wafers were subjected to a high voltage for about 1 hour. This voltage (poling voltage) produces a remanent polarization in an initially disordered unpolarized ceramic. Once poled, a wafer need not be repoled.

Measurements were also made of a photovoltaic output from two single crystals of barium titanate. These open-circuit photovoltages at room temperature are shown in Table II. At band gap, the 0.2-cm - thick crystal produced a short circuit photocurrent of $.006 \mu\text{A}/\text{cm}/\mu\text{W}/\text{cm}^2$.

Ceramic materials produce much higher voltages proportional to length, whereas the single crystals produce low voltages and certainly not proportional to length. These differences between the polycrystalline ceramic material and the single-crystal material suggest to us that the high voltages across the polycrystalline wafers result from the series addition of many small voltages across individual crystallites of the ceramic. The crystallite voltage would be roughly that seen across the single crystal, about 0.3 V. Barium titanate ceramic produced about 300 V/cm implying that the barium titanate grains must be about 10μ across, which is roughly correct. Further verification of the additive nature of individual grain photovoltage is found in the open-circuit voltage results for PLZT ceramics in Table I. The photovoltage across material made of 2 to 4μ grains is roughly twice that of the identical material made up of roughly 6μ grains.

For these net voltages to appear, a unique axis and direction must be defined in the individual crystallites, and there must be an alignment of these. The unique directed axis for a ferroelectric would reasonably be the direction of polarization, with the photo-induced field lying parallel to the crystal polarization axis (the c-axis). The net alignment in a polycrystalline ceramic would then be produced by the poling process in which the application of a large DC voltage orders the directions of polarization domains, producing a net remanent polarization in originally disordered ceramic material. The photovoltage in the ceramic material would then vanish for a disordered material and be a maximum for ordered material. This is illustrated in the experimental results shown in Fig. 4. There is, in fact, a linear relation between remanent polarization and photovoltage. The remanent polarization was varied by applying different poling voltages. The photovoltage was then measured. The net remanent polarization was determined afterwards by measuring the charge released when the wafer was thermally depoled.

Since the individual crystallite photovoltage seems somehow related to the direction and orientation of the polarization the photovoltages across the crystallites and the photovoltage across the macroscopic single crystal must vanish when the crystallites and the macroscopic crystal are in a centrosymmetric state. The transition between the tetragonal polar ferroelectric state and the cubic centrosymmetric non-polar paraelectric state occurs in barium titanate at about 120°C (roughly also the Curie temperature). The photovoltages (V_0), in both the single and polycrystalline material do, as expected, vanish there as shown in Fig. 5 and Fig. 6.

The remanent polarization of a wafer was obtained as a function of temperature by measuring the pyroelectric charge released as the temperature was raised. It is shown (marked P_r) in Fig. 6. The polarization vs T is also shown for the single crystal (Ref. 9).

To assure ourselves that this was a true photovoltaic effect in which charge carriers were injected into the conduction band, absorbing energy from incident photons, we examined the spectral dependence of the short-circuit photocurrent (Fig. 7, 8, 9 and 10). The photocurrent peaks, as expected, in the vicinity of band-gap radiation (Ref. 8). The intensities available from the apparatus varied somewhat as filters were changed. Short circuit current is, however, proportional to intensity (Fig. 12) allowing results to be normalized to constant intensity. Also, as might be expected, there is a spectral dependence for the open-circuit voltage; this is shown for PZT in Fig. 11. Intensity varied somewhat with wavelength. The results were, however, obtained in a saturated region where open-circuit voltage was not a strong function of intensity (Fig. 12).

The relation between short-circuit current and intensity, and open-circuit voltage and intensity are both shown in Fig. 12. The material is PZT. The short-circuit current is linear with intensity. The open-circuit voltage is linear with intensity for low intensities, saturating at higher levels. Short-circuit photocurrent was also measured as a function of temperature and the results are shown in Fig. 13 for $\text{BaTiO}_3 + 5 \text{ wt.}\% \text{ CaTiO}_3$ and in Fig. 14 for PZT.

Further experimental measurements are obviously desirable for a better understanding of this high-voltage photovoltaic effect in ferroelectrics. Our experience now indicates that all ferroelectric oxide perovskites are good candidates for exhibiting the effect.

APPLICATIONS

This is a new effect with possible applications. Clearly, this effect could be used to produce very high voltages in a relatively uncomplicated fashion. Of greater importance, perhaps, is a possible potential for direct solar-to-electric energy conversion, although this would necessitate finding a material with considerably greater photovoltaic efficiencies over broader spectral regions.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the help of staff members, particularly James Blackburn for helpful discussions and his assistance in developing instrumentation; also Stanley Watkins who made many of the measurements and helped in assembling this paper.

REFERENCES

1. P.S. Brody, Solid State Communications, 12 pp. 673-676, 1973.
2. Brody, P.S. and Vrabel, M.J. Bull. Am. Phys. Soc. 13, 617 (1968).
3. Chynoweth, A.G., Phys. Rev. 102, 705 (1956).
4. Uchida, N. and Ikeda, T., Japan J. Appl. Phys. 7, 1219 (1968).
5. For example see, Pensak, L., Phys. Rev. 109, 601 (1958).
6. For a review see, Tauc, J., Photo and Thermoelectric Effect in Semiconductor, Pergamon Press, London, (1962).
7. Merz, W., Helv. Phys. Acta 31, 625 (1958).
8. Cox, G.A., Roberts, G.G. and Tredgold, R.H., Br. J. Appl. Phys. 17, 793 (1966).
9. Merz, W.J., Phys. Rev. 91, 513 (1953).

Table 1. Electrical Outputs for Several Ceramics with Various Illuminations
(The PZT materials contain additional additives of about 1 wt%)

Sample Material	Number of Samples Tested	Wavelength (Micrometers)	One-Half Bandwidth (Micrometers)	Intensity ($\mu\text{watts/cm}^2$)	Open Circuit Voltage (Volts/cm)	Short Circuit Current ($\mu\text{amps/cm}$)	Photovoltaic Efficiency (%)
PZT 53/47 Source A	1	373	7	100	610	.31	.019
PZT 53/47 Source A	4	373	7	1000	430-30	.36	.015-0.002
PZT 53/47 Source B	1	373	7	1000	580	.20	.012
PZT 53/47 Source C	1	373	7	1000	745	.22	.016
BaTiO ₃ + 5wt% CaTiO ₃	1	403	7	3500	355	.019	.0007
PLZT 7/65/35 2-4 μ grains 1	1	382	7	-	1120	.03	.0034
PLZT 7/65/35 >6 μ grains 1	1	382	7	-	630	.04	.0025
PLZT 8/65/35 2-4 μ grains 1	1	382	7	-	500	.012	.0006
PLZT 8/65/35 >6 μ grains 1	1	382	7	-	300	.02	.0006
PZT 53/47 Source A	8*	370	80	1815	440-20	.085-0.04	.005-0.0004
PZT 53/47	8**	SUNLIGHT	-	80,000	-	-	.0004

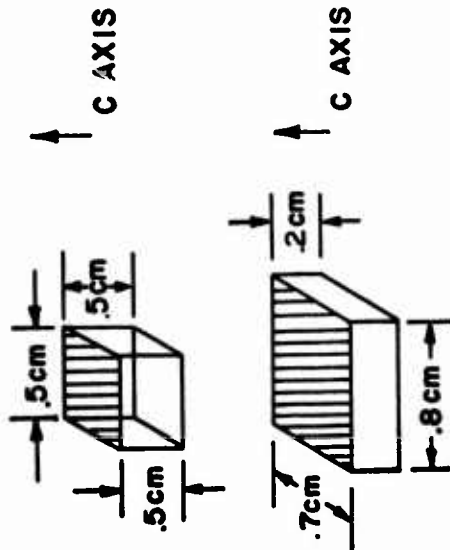
*includes 4 samples of line 2 of the table

**A series parallel combination of 8 2.3 x 11.8 cm wafers yielding an open circuit voltage of 4300 volts.

TABLE II: Open-circuit Voltage Measurement for Barium Titanate Single Crystals (Electrodes were perpendicular to the c-axis)

Open-circuit Voltage
(V_o)

Crystal Dimensions



.25

.55

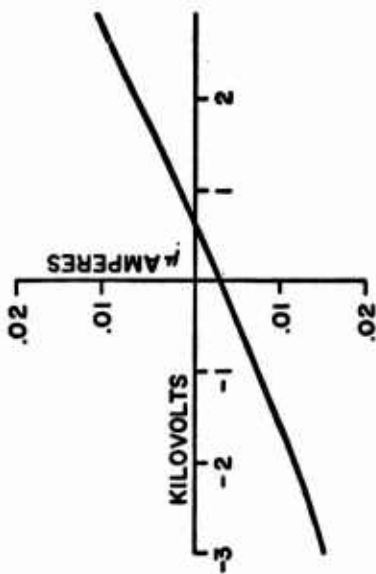


FIG. 2 CURRENT vs APPLIED VOLTAGE, ILLUMINATED PZT WAFER

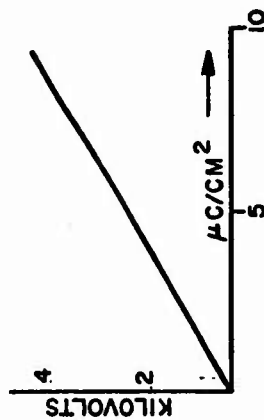


FIG. 4 PHOTOVOLTAGE vs REMANENT POLARIZATION, $\text{BaTiO}_3 + 5\text{wt}\% \text{CoTiO}_3$ CERAMIC

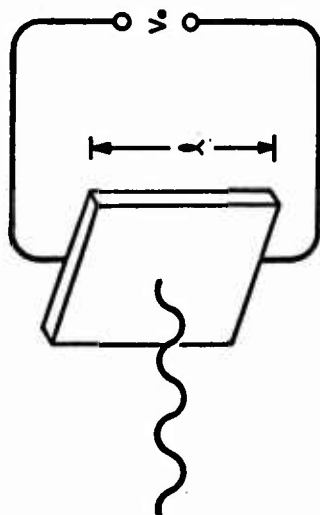


FIG. 1 ILLUMINATED WAFER.

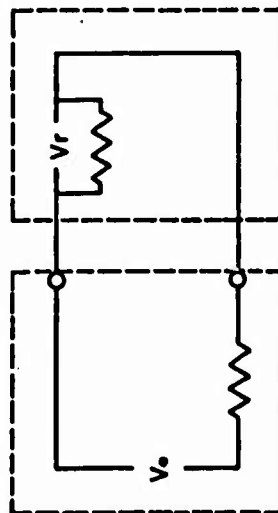


FIG. 3 WAFERS CONNECTED TO EXTERNAL CIRCUIT

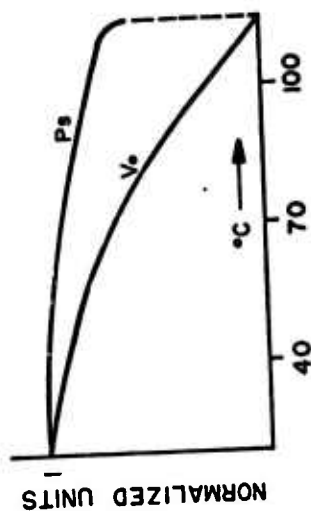


FIG. 6 PHOTOVOLTAGE vs TEMPERATURE, BaTiO_3 SINGLE CRYSTAL

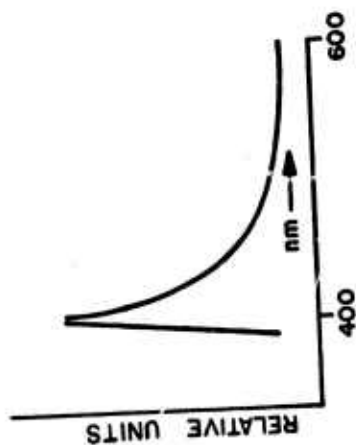


FIG. 8 PHOTOCURRENT vs WAVELENGTH, SINGLE CRYSTAL BaTiO_3

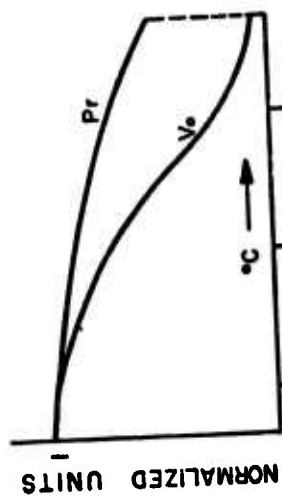


FIG. 5 PHOTOVOLTAGE vs TEMPERATURE, $\text{BaTiO}_3 + 5\text{wt}\% \text{CaTiO}_3$

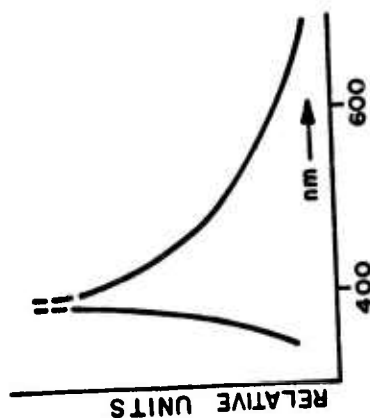


FIG. 7 PHOTOCURRENT vs WAVELENGTH, $\text{BaTiO}_3 + 5\text{wt}\% \text{CaTiO}_3$

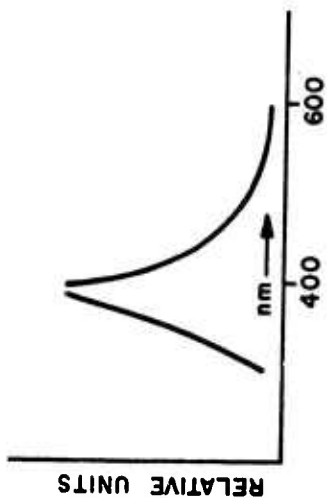


FIG. 9 PHOTOCURRENT vs WAVELENGTH, PZT 53/47

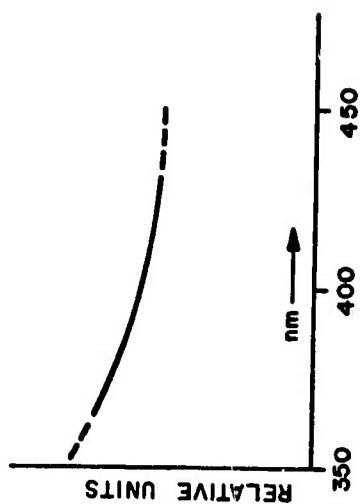


FIG. 11 PHOTOVOLTAGE vs WAVELENGTH, PZT 53/47

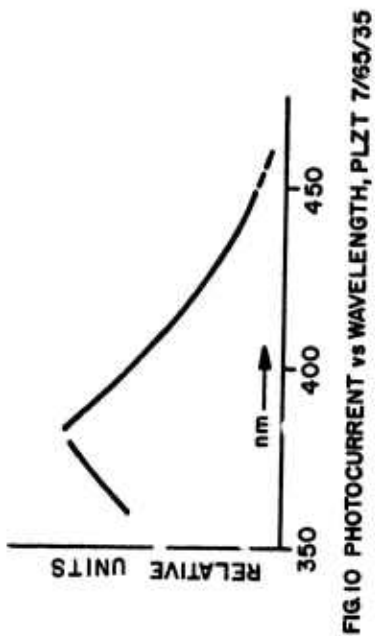


FIG. 10 PHOTOCURRENT vs WAVELENGTH, PLZT 7/65/35

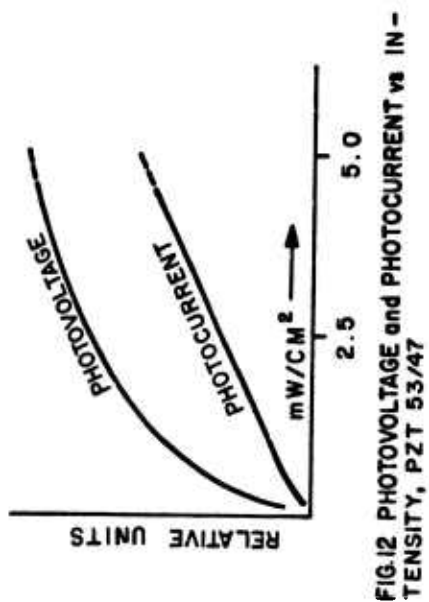


FIG. 12 PHOTOVOLTAGE and PHOTOCURRENT vs INTENSITY, PZT 53/47

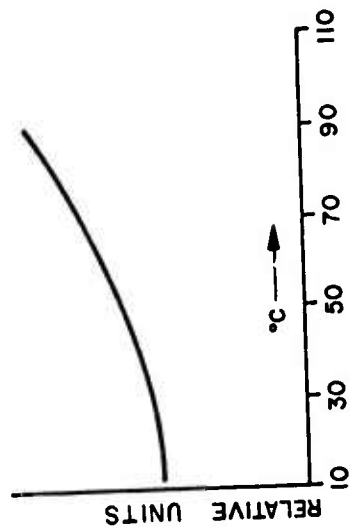


FIG. 14 PHOTOCURRENT vs TEMPERATURE, PZT 53/47

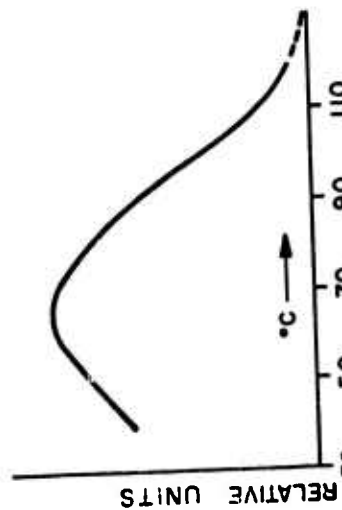


FIG. 13 PHOTOCURRENT vs TEMPERATURE, BaTiO₃

OCULAR AND SKIN HAZARDS FROM CO₂ LASER RADIATION

ARNOLD S. BROWNELL, DR.
BRUCE E. STUCK, MR.*
JOINT AMRDC-AMC LASER SAFETY TEAM
FRANKFORD ARSENAL, PHILADELPHIA, PENNSYLVANIA

INTRODUCTION

The rapid development of new laser systems with greater military significance continues unabated. The current and projected field deployment of these systems has intensified the need not only to define the potential hazards to military personnel but also to investigate the mechanisms involved in laser induced radiation injuries. For example, many of the newer carbon dioxide (CO₂) lasers are potentially hazardous because of their high power output. Definitive data are urgently required to provide additional information essential to the establishment of safety standards which would be relevant to all laser systems, for use by both system developers and field personnel.

Research establishing laser damage thresholds have generally been empirical. Each present and future generation laser system has or will have unique and varied radiation parameters such as wavelength, intensity, repetition rate and beam geometry. Similarly, the various vulnerable tissues have their own distinct physical, chemical and biological properties. Expensive and time consuming laboratory experiments are required to estimate threshold damage levels for all possible laser systems. Consequently, there is a need for a more expeditious approach. Mathematical models that accurately predict the extent of laser induced damage and provide insight into the mechanisms of the interaction of laser radiation with biological systems will significantly improve cost effectiveness.

The purpose of this paper is threefold: First, to provide data necessary for military and civilian safety communities by

Preceding page blank

presenting experimentally determined threshold doses, from two independent studies, for the minimal detectable changes in cornea and skin following exposure to CO₂ laser radiation. Second, to test the validity of a mathematical model in predicting damage thresholds. Third, to determine the extent this model accounts for differences in the experimentally determined dose-response relationships for the two tissues studied.

EXPERIMENTAL METHODS

Details of the experimental procedures, and results, have been reported (1-5). A Coherent Radiation Model 41 continuous wave CO₂ laser was used for skin exposures of less than 0.1 sec in duration and for all corneal exposures. For these experiments the beam incident on the tissue was formed by illuminating a circular aperture with the laser beam. The radiant intensity distribution in the central portion of the resulting diffraction pattern closely approximated a Gaussian distribution. The relative intensity distribution was measured with a scanning thermal sensor. The total power was measured with a CRL Model 201 power meter whose calibration had been checked at the National Bureau of Standards. The values reported in the paper represent the peak irradiance (watts/cm²), calculated on the basis of the measured total power of the beam and its relative intensity distribution. For skin exposures in excess of 0.1 sec, a CO₂ laser designed and constructed by Martin Marietta Corporation provided the radiant energy. The multimodal, collimated output beam consisted of a mosaic of many closely spaced regions with different intensities. The relative intensity distribution within the beam varied within 25%. The reported values represent the average irradiance in the beam.

Rabbits (*Oryctolagus Cuniculus*) and monkeys (*Macaca Mulatta* and *Aotus Trivirgatus*) were used in the cornea threshold studies. Previous results (3) by this laboratory have shown less than a 10% difference between the damage thresholds for these three animal species. Pentobarbital sodium and halothane gas were used to anesthetize the monkeys and rabbits respectively. The pupils were dilated to enhance the observer's visualization of corneal changes. The eye was gently irrigated with physiological saline before exposure and closed for approximately 1 minute to prevent drying of the corneal surface and to maintain a near "natural" tear film. The lid was retracted and immediately irradiated at normal incidence. Two to four exposure sites on each eye of fifteen to twenty animals were irradiated for each threshold determination. The criterion for a minimally detectable response was the appearance of a greyish-white circular opacity

on the outer surface of the cornea as seen with a Zeiss biomicroscope 30 minutes post exposure.

For skin experiments, white pigs, 20 to 40 pounds each, were the test animals. Pentobarbital sodium was used as an anesthetic for animals whose skin was exposed in excess of 0.1 sec. Those whose exposure times were less than 0.1 sec received halothane gas. Three to eight animals were used for each irradiance level. The skin test area was closely clipped, cleaned, and then divided into a grid pattern of 4 rows and 10 columns. To minimize variation in area-to-area sensitivity, the exposure times were randomly assigned to the grid pattern of each animal within each irradiance group. The visual appearance of each test site was observed at the time of exposure and within 24 hours post exposure. The criterion for a positive response for a minimal lesion was the presence of a mild erythema 24 hours post exposure.

EXPERIMENTAL RESULTS

The probit method (6,7) was used in the analysis of the data. This statistical technique was developed to treat quantal response data and define dose-effect relationships. This analysis yields the median effective dose (ED_{50}), which is the irradiance for a given exposure time that has a 50% probability of producing the specified response. In this paper the ED_{50} value is defined as the damage threshold for a specified injury.

Figure 1 presents a log-log plot of the experimentally determined ED_{50} values for both CO_2 laser radiation induced minimal erythema and corneal opacities. The exposure durations range from 1 msec to 7.4 seconds while the irradiances vary from 1.2 to 800 watts/cm². The numerical values represented by the data points are used to generate two of the variables in the model described in the following section.

DAMAGE INTEGRAL MODEL

Theoretical Development

Most injuries resulting from exposure to laser radiation are considered to be thermal in nature. The degradation of the radiant laser energy to thermal energy leads to temperature rises in the exposed tissue. The basic principles involved in the analysis of thermal injuries in tissue were defined in a series of papers by Henriques and Moritz. Henriques (8) considered the development of thermal injuries to be time and temperature dependent, a rate process

common to many physical and biological systems. He suggested that by integrating the injury rate process over the thermal time period involved, the contributions from each time increment could be summed to yield a measure of the total damage incurred.

If the kinetics of any specific state of thermal injury can be regarded as a first order reaction at all temperatures under consideration, then the sum of the injury can be expressed as

$$\Omega = \int_{t_0}^t k dt \quad (1)$$

where k is the specific reaction rate constant for a given thermal process and temperature, and Ω is the damage integral.

If the reaction is similar to that of many chemical and physical rate processes, the thermal inactivation rate constant should vary with temperatures as

$$k = Ae^{-B/T} \quad (2)$$

when A and B are constants and T is the absolute temperature. For a reaction meeting the necessary thermodynamic criteria, this expression is equivalent to the integrated form of the Arrhenius equation, where B equals $\Delta E/R$, ΔE being the energy of activation and R the gas constant. In the present use of the model, A and B are considered empirical constants with no special implications.

Combining Equations 1 and 2 yields

$$\Omega = A \int_{t_0}^t e^{-B/[T(t)+T(0)+273]} dt \quad (3)$$

where $T(t)$ is the temperature rise at time t and $T(0)$ is the initial tissue temperature in $^{\circ}\text{C}$. This concept is equivalent to that proposed by Henriques (8) and subsequently expanded by Fugitt (9).

A mathematical equation that defines $T(t)$ during the thermal episode for the various exposure and tissue conditions is required by Equation 3. The diathermanous, semi-infinite heat flow equation given by Carslaw and Jaeger (10) and extended by Davis (11) to apply to problems involving radiant heating of tissues is appropriate in

situations where the heat flow in the tissue space of concern can be considered to be unidirectional and perpendicular to the surface. This equation has the following form:

$$T(t) = \frac{2H}{\sqrt{\mu}} \left\{ \sqrt{\frac{t}{\pi}} e^{-l^2/4\alpha t} - \frac{l}{\sqrt{4\alpha}} \operatorname{erfc}\left(\frac{l}{\sqrt{4\alpha t}}\right) - \frac{e^{-\gamma l}}{\gamma\sqrt{4\alpha}} + \frac{e^{\gamma^2 2\alpha t}}{4\sqrt{\alpha}\gamma} \left[e^{\gamma l} \operatorname{erfc}\left(\gamma\sqrt{\alpha t} + \frac{l}{\sqrt{4\alpha t}}\right) + e^{-\gamma l} \operatorname{erfc}\left(\gamma\sqrt{\alpha t} - \frac{l}{\sqrt{4\alpha t}}\right) \right] \right\}. \quad (4)$$

Although the diathermanous heat flow equation assumes a radiation beam infinitely wide with no variation in its intensity distribution, the beam widths were sufficiently large for all skin exposures to justify the use of this equation in these calculations.

Where radial heat flow must be taken into consideration, Heimbach's combined model (12) is more applicable, as follows:

$$T(t) = e^{-\gamma l} \sum_i \frac{\delta_i I_0(r_i \frac{t}{d})}{\gamma^2 - (\frac{r_i}{d})^2} \left\{ e^{\alpha \gamma^2 t} e^{-\alpha (\frac{r_i}{d})^2 t} - 1 \right\} + \sum_i \frac{\gamma \delta_i I_0(r_i \frac{t}{d})}{\gamma^2 - (\frac{r_i}{d})^2} \left\{ \frac{1}{2(\frac{r_i}{d})} \left[e^{-r_i l/2} \operatorname{erfc}\left(\frac{l}{\sqrt{4\alpha t}} - \sqrt{\alpha} \left(\frac{r_i}{d}\right) \sqrt{t}\right) - e^{-r_i l/2} \operatorname{erfc}\left(\frac{l}{\sqrt{4\alpha t}} + \sqrt{\alpha} \left(\frac{r_i}{d}\right) \sqrt{t}\right) \right] - \frac{1}{2\gamma} e^{\alpha \gamma^2 t} e^{-\alpha (\frac{r_i}{d})^2 t} \left[e^{-\gamma l} \operatorname{erfc}\left(\frac{l}{\sqrt{4\alpha t}} - \sqrt{\alpha \gamma^2 t}\right) - e^{\gamma l} \operatorname{erfc}\left(\frac{l}{\sqrt{4\alpha t}} + \sqrt{\alpha \gamma^2 t}\right) \right] \right\} \quad (5)$$

This equation considers a radiation beam finite in cross section with an intensity distribution exhibiting cylindrical symmetry. It was used to fit the corneal data where the small beams require the more complex equation.

Equations 4 and 5 define the temperature rise only during the exposure interval. During the post exposure period the complete expression is the identical equation minus the same equation where t is replaced by $t - \eta$ (η = exposure time). The complete development of these equations is given by Heimbach (12).

The following assumptions are applied to the application of both heat flow equations: the tissue is a semi-infinite isotropic receiver, initially at a uniform temperature throughout; its thermal properties are constant and do not vary with temperature; its surface is perfectly insulated and there are no radiation losses; the radiation input is a square wave and normal to the surface and the tissue absorbs the radiation at an exponential rate.

Thus Equations 4 and 5 are used to calculate $T(t)$ in Equation 3 to obtain the damage integral Ω . To determine the constants A and B (Equation 3) their numerical values were adjusted until the calculated dose yielding an $\Omega = 1$ corresponded closely to each experimentally determined ED_{50} value for the minimal responses given in Figure 1.

Figure 2 shows representative histological sections of normal pig skin and monkey cornea. Although morphologically the two tissues are complex, they can be structurally described as simple three layer systems.

In the skin the three layers are: (1) The outer layer, the stratum corneum, is composed of compacted keratinized scales. The average thickness of this nonviable layer is approximately 15 microns. (2) The more deeply stained layer is the viable portion of the epidermis. Its thickness is quite variable, but in general ranges from 50 to 75 microns. (3) The underlying dermis, composed primarily of a matrix of collagen and elastic fibers, ranges up to several hundred microns in depth. Capillary loops of the skin blood supply and many nerve endings terminate within the upper portion of this layer.

Similarly the corneal layers are: (1) In vivo a 7-9 micron thick tear film covers the cornea (13). Obviously this tear film is not evident in Figure 2. (2) The epithelium averages 40 microns in thickness and, as in the skin, is capable of regeneration. (3) The stroma comprises about 90% of the entire cornea and is approximately 400 microns thick. It consists of collagenous lamellae.

The damage integral model, as presented, assumes that the thermal episode occurs in an isotropic medium. It is evident that neither tissue is homogeneous in a morphological sense. However, with the exception of the cornified upper layer of the skin, the various layers of the tissue are of an aqueous nature with a water content of approximately 80%. So, in a physical sense, as a first approximation they can be considered homogeneous and satisfy the requirements of the model.

Table I defines the variables used in the temperature calculations. The last 2 columns give the numerical value of those variables held constant to determine the damage integrals.

TABLE I

		<u>Cornea</u>	<u>Skin</u>
μ	Thermal Inertia (cal ² cm ⁻⁴ deg ⁻² cm ⁻¹)	1.5×10^{-3}	1.18×10^{-3}
α	Thermal Diffusivity (cm ² sec ⁻¹)	1.5×10^{-3}	0.86×10^{-3}
γ	Absorption Coefficient (cm ⁻¹)	814	814
T(0)	Initial Tissue Temperature (°C)	32	35
r	Distance from Beam Center (cm)	0	0
z	Tissue Depth (cm)	.0010	.0025 .0100
H	Irradiance (cal sec ⁻¹ cm ⁻²)		
t	Time After the Onset of the Exposure (sec)		

The complex variable δ_1 is defined in the following equation:

$$\delta_1 = \frac{\sqrt{\alpha\mu} A_1}{\gamma H}$$

where A_1 are the expansion coefficients obtained by expanding the intensity distribution of the radiation beam into a Fourier-Bessel series of zeroth order. The roots of the zeroth order Bessel function are represented as r_1 and 'a' is the radial extent of the expansion. For the corneal calculations the intensity distribution is considered Gaussian with a 1/e radius of 0.14 cm.

In the case of the skin, the two thermal constants, μ and α , were experimentally determined by a radiant heating method (11) for porcine skin. Measurements of the absorption of 10.6 micron radiation by epidermal tissue (14) was not significantly different from that of water; therefore, the latter absorption coefficient (15) was used. T_0 was chosen as a close approximation to a measured mean temperature of 35.2°C for 3 sites on each of 12 experimental animals.

Since the cornea including the tear film is essentially aqueous in nature, the thermal constants and the absorption coefficient of water were chosen (16,15). For the initial corneal temperature the experimentally measured value of 32°C was used (17).

For the cornea, the minimal response is direct and immediate. The resulting opacities were 200 microns in diameter. The fact that the coagulum disappears within 24 hours substantiates the suggestion that only the anterior portion of the epithelial layer is involved. Consequently, a depth of 10 microns, which includes the tear film, was selected to calculate the damage thresholds for this tissue.

On the other hand, the erythemic reaction is a secondary response. Epithelial injury produces a substance which diffuses to the underlying capillary bed and thus stimulates an increase in blood volume in that region. The result is an increase in redness relative to the surrounding tissues, as seen through the translucent epidermis. The minimum histological changes in tissue samples showing an erythemic response were detected within a 10 micron depth in the viable layer of the epidermis. The depth of 25 microns chosen to calculate damage thresholds for the erythemic reaction includes the 15 micron cornified layer of the epidermis.

Results and Discussion

The curves in Figure 1 represent the values generated by the damage integral model for the minimal responses. The inactivation coefficients (from Equation 3) were obtained by fitting the curves to the experimentally derived data points given in the figure. They are:

	<u>Skin</u>	<u>Cornea</u>
ln A	146.5	141.6
B	49,000	48,400

The nearly equivalent numerical values for B suggest that the two responses have essentially the same inactivation energy requirements.

The two curves in Figure 1 have some similarity in shape. At the lower exposure times both show an increase in slope with decreasing exposure time. However for pulse durations in excess of 0.1 sec, while the slope for the erythemic curve is constant, that for the corneal opacity data shows a significant change.

The same curve fitting the corneal opacity data for the time interval from 0.1 to 40 sec is given in Figure 3. The calculations were repeated, using the same input parameters with the diathermanous heat flow equation which generated the second dashed line (- - -). The resultant curve also shows a constant slope for the longer exposure times. For times less than 0.1 sec the calculated damage threshold values are essentially identical using either heat flow equation. Therefore the difference in the two curves can only be attributed to the radial heat flow in the tissue.

This difference reemphasizes the need to carefully consider beam geometry in deriving safety standards from experimental data. The model predicts even at depths less than 10 microns the beam size can significantly affect the amplitude and duration of the thermal episode. This phenomenon markedly increases with increasing depth.

As a further test of the predictive value of the model, damage thresholds for white burns in skin were calculated and compared with experimentally derived values. White burns in the skin are similar to the corneal opacities. They are an immediate coagulation of the superficial layers of the skin. However, the opacification must be rather extensive in order to produce enough contrast to detect this level of damage. Hysell, et al (18), indicated that white skin burns had histopathological effects extending into the dermis, some of which showed coagulated collagen bundles in the superficial dermis. For calculations involving these burns, a depth of at least 100 microns appears appropriate.

Using the inactivation coefficients obtained for the corneal data, a depth of 100 microns, and all other parameters used for the skin, the solid curve in Figure 3 was computed using the diathermanous heat flow equation. The solid data points represent the experimentally determined ED₅₀ values for 'white' burns in porcine skin. The data points reasonably correspond to the theoretically derived curve except for exposures of 1 sec and less where the experimental results deviate significantly from those predicted. At these exposure times, the calculated temperature rise at the surface of the aqueous portion of the epidermis exceeds 100°C. Thus a portion of the absorbed radiant energy must be expended into the latent heat of vaporization. Consequently, more incident energy would be required to sufficiently

elevate the temperature at a 100 micron depth. This possibility is supported by experimental results using shorter exposure times in which steam blebs appeared during the exposure episode at doses lower than required to produce detectable white burns. The present model requires further expansion in order to account for these responses.

Rigorous testing of this model, which has been expanded to include cases of tissue damage from repetitively pulsed laser beams, as well as a two layer model is continuing.

From the foregoing discussion it can be generally concluded that: (1) The damage integral model can be used to predict the minimal response of the skin and cornea to CO₂ laser radiation. (2) The model describing the dose-response relationship for the coagulative process in the cornea gives a reasonable prediction for a similar response mechanism, white burn, in porcine skin within the limits assumed by the model. (3) Confirmation of the inactivation coefficients for the corneal and skin response mechanisms will permit the model to be used in predicting damage thresholds for different laser systems. (4) The success of this model over such a wide range of exposure durations and irradiances justifies its further testing and possible expansion in order to realize a significant reduction in the time and money required to provide data relevant to the evaluation of laser radiation hazards.

Acknowledgements

Technical computer assistance by Sp 5 Frederick L. Buchta is gratefully appreciated. Histopathological support was provided by Dr. Peter J. Ihrke, VMD, School of Veterinary Medicine, and Dr. Myron Yanoff, MD, School of Medicine, University of Pennsylvania.

In conducting the research described in this report, the investigators adhered to the "Guide for Laboratory Animal Facilities and Care" as promulgated by the Committee on the Guide for Laboratory Animal Facilities and Care of the Institute of Laboratory Animal Resources, National Academy of Sciences - National Research Council.

REFERENCES

1. Brownell, A.S., W.H. Parr, D.K. Hysell and R.S. Dedrick: Thresh-
hold lesions induced in porcine skin by CO₂ laser radiation.
USAMRL Report No. 733, 1967 (DDC AD-659 347).
2. Brownell, A.S., W.H. Parr, D.K. Hysell and R.S. Dedrick: CO₂
laser induced skin lesions. USAMRL Report No. 769, 1968
(DDC AD-669 610).
3. Byer, H.H., E. Carpino, B.E. Stuck: Determination of the thresh-
holds of CO₂ laser corneal damage to Owl monkeys, Rhesus monkeys
and Dutch Belted rabbits. Frankford Arsenal Report M72-31,
Feb. 1972 (DDC AD-901 0862).
4. Stuck, B.E., E. Carpino: Ocular effects of CO₂ laser radiation
for pulse durations from 1 msec to 5 sec, in preparation.
5. Brownell, A.S., R.J. Rodgers, Jr., T.W. Elverson and F.L. Buchta:
Damage thresholds for porcine skin exposed to millisecond pulses
of CO₂ laser radiation, in preparation.
6. Finney, D.J.: Probit Analysis, Cambridge, Mass., University
Press, 1952.
7. Litchfield, J.T., Jr., and Wilcoxon: A simplified method of
evaluating dose effect experiments. J. Pharm. Exp. Therap.
96:99, 1949.
8. Henriques, F.C., Jr.: Studies of thermal injury v. The pre-
dictability and the significance of thermally induced rate
processes leading to irreversible epidermal injury. AMA Arch.
Path. 43:489, 1947.
9. Fugitt, C.H.: A rate process theory of thermal injury. AF Spe-
cial Weapons Project Report No. 606, 1955 (DDC AD-212 660).
10. Carslaw, H.S. and J.C. Jaeger: Conduction of Heat in Solids,
Sec. Ed., Oxford, Carenden Press, 1959.
11. Davis, T.P.: A theoretical and experimental investigation of
the temperature response of pig skin exposed to thermal
radiation. Univ. of Rochester Atomic Energy Report No. UR-533,
1959.

12. Heimbach, C.R.: Models of heat dissipation in a homogeneous solid. USAMRL Report No. 954, 1971 (DDC AD-738 789).
13. Mishima, S.: Some physiological aspects of the precorneal tear film. Arch. Ophthalm. 73:234, 1965.
14. Hardy, J.D. and C. Muschenheim: Radiation of heat from the human body v. The transmission of infra-red radiation through skin. J. Clin. Invest. 15:1, 1936.
15. Robertson, C.W., D. Williams: Lambert absorption coefficients of water in the infrared. J. Opt. Soc. Am. 61:1319, 1971.
16. Weast, R.C. ed., CRC Handbook of Chemistry and Physics, 48th edition.
17. Schwartz, B. and Fellar, M.R.: Temperature gradients in the rabbit eye. Inv. Ophthalm. 1:516, 1962.
18. Hysell, D.K., and A.S. Brownell: Correlation between the gross and microscopic appearance of CO₂ laser induced porcine skin burns. USAMRL Report No. 778 (DDC AD-676 578).

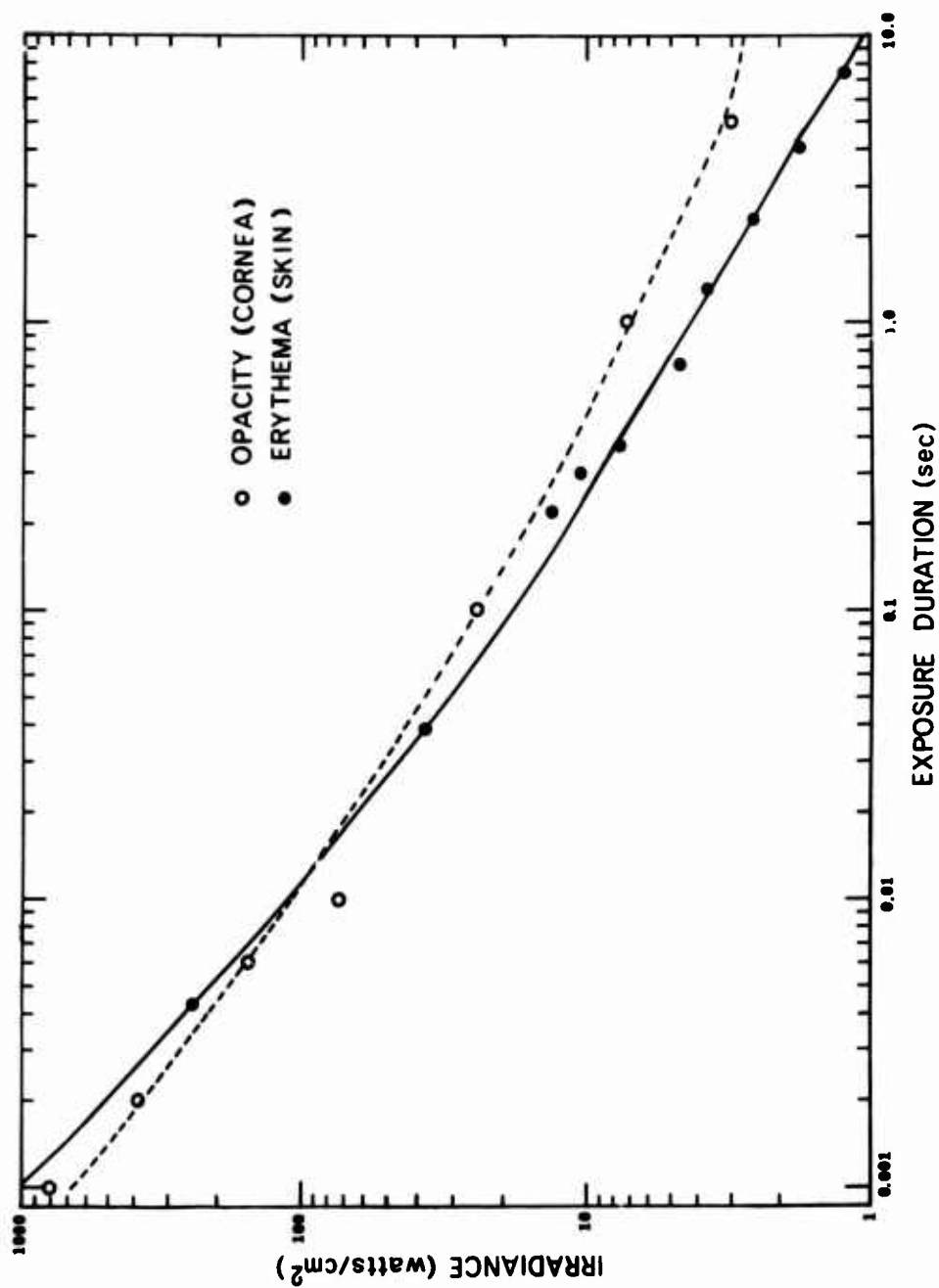


Figure 1. Experimentally derived ED_{50} values for CO_2 laser radiation induced minimal erythema in skin and corneal opacities.

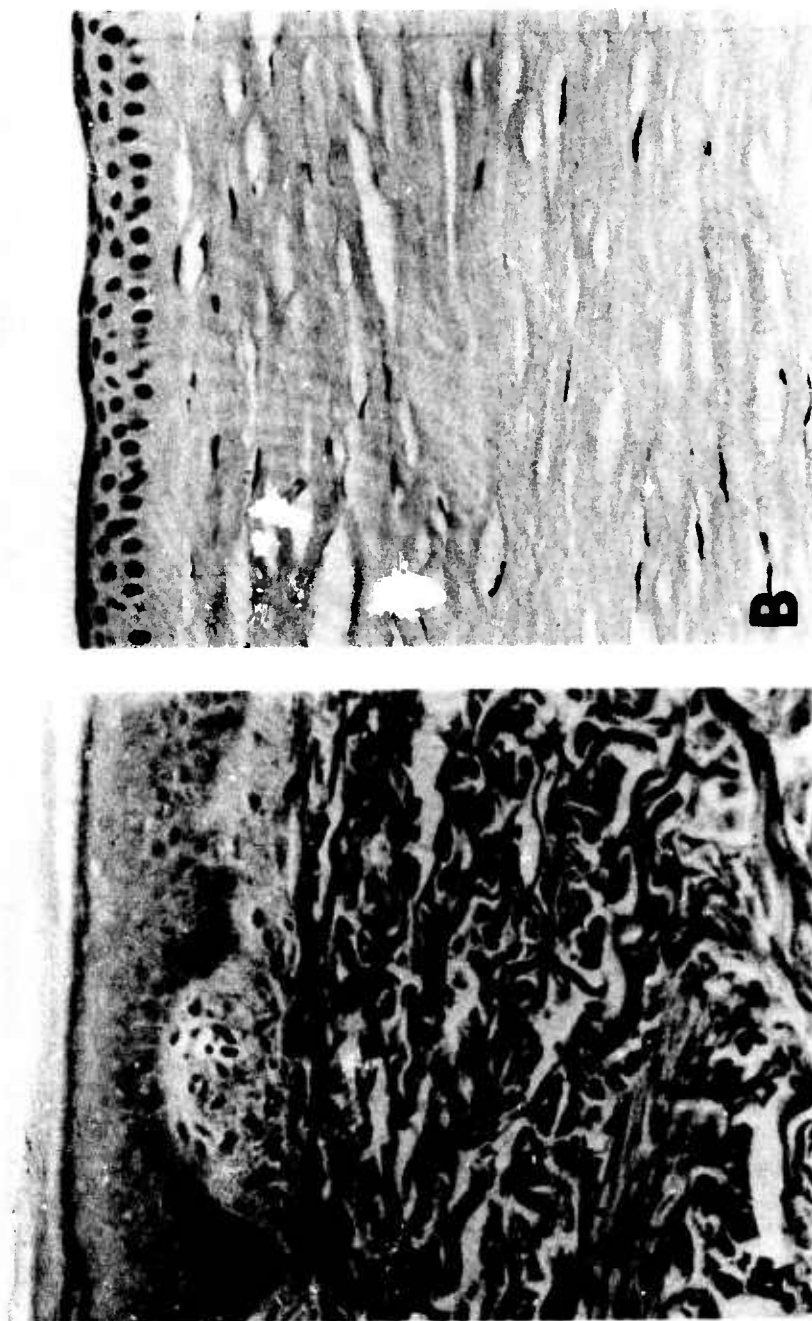


Figure 2. Histological sections of normal porcine skin (A) and monkey cornea (B) stained with hematoxylin and eosin. Separation of skin corneum is artifactual.

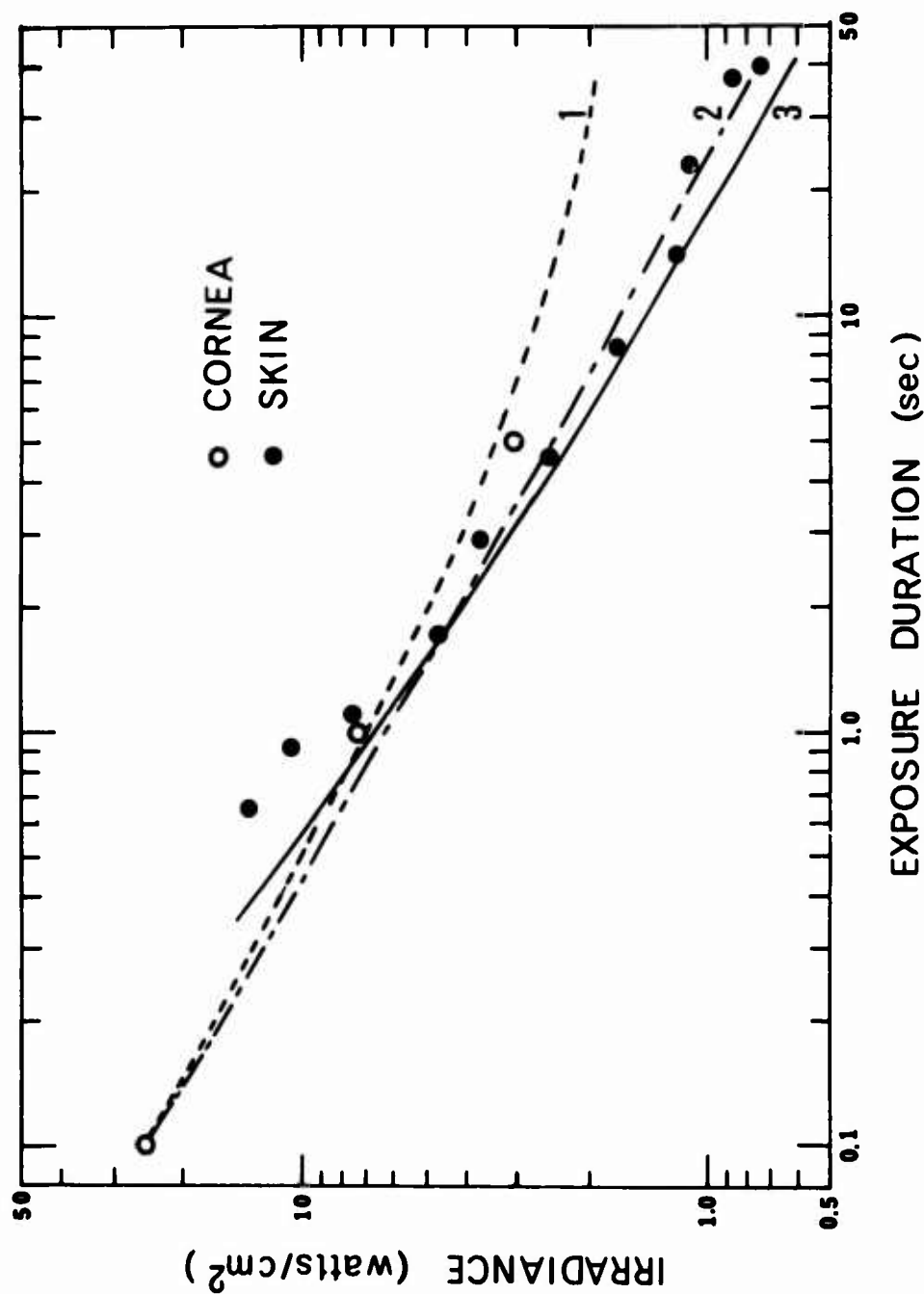


Figure 3. Comparison of coagulative effects using the damage integral model. Combined model for corneal opacities (1); diathermanous model for corneal opacities (2); and white skin burns (3).

VAPORIZATION WAVES IN IMPULSIVELY HEATED MATERIALS

HENRY S. BURDEN, MR.

U. S. ARMY BALLISTIC RESEARCH LABORATORIES
ABERDEEN PROVING GROUND, MARYLAND

1. INTRODUCTION

Computer hydrocodes are essential in determining the magnitude, transmission, and effects of stresses produced when energy is impulsively deposited in a material. As do any solutions of the hydrodynamic equations, these numerical approaches depend upon the form of an auxiliary equation of state for their success. Most hydrocodes give satisfactory treatment in the solid and liquid phases but in codes (1) such as PUFF and RIP the equations of state ignore or improperly treat the liquid-vapor phase change region in which the major expansion of subcritically heated materials takes place. In this paper, a model for this expansion is described which was successful in resolving an anomaly in exploding wire response (2) but is of a type that could not be duplicated by hydrocodes using these equations of state. With refinement, this vaporization wave model can be applied to determining thermodynamic data that will aid in revising the equations of state for materials of interest. These improvements will have relevance to the response of metals impulsively converted to vapor by absorption of radiation, whether it be from laser, electron beam, nuclear, or other sources.

2. IMPULSIVE HEATING

If a substance is so rapidly heated that the normally associated expansion is prevented by the material's own inertia, pressure will be built up in its interior and maintained at any point until an expansion wave originating at a free surface has had time to penetrate to that point. A representative P - ρ diagram (3) (Figure 1)

Preceding page blank

may be used to describe the expansion following such heating. To place it in our realm of interest, we will choose a process which causes the material to enter the liquid vapor coexistence state. Such a process may be imagined as follows: the point describing the state of a material will, if the material is heated instantaneously, jump from its room temperature, uncompressed position vertically until it reaches a new position representing an unchanged density and lying on a contour marked with the value of heat energy added, e.g., $E=2000$ cal/g. This would correspond on the diagram of Figure 1 to a pressure on the order of 400 kbar ($1 \text{ bar} = 10^6 \text{ dyne/cm}^2 \cong 1 \text{ atm}$). Then, if the expansion takes place over a period less than a few microseconds, negligible heat will be lost from the material by radiation or conduction and we may follow the energy contour (adiabat) as the material expands to lower values of density. However, material at no point except on a free surface may expand unless adjacent material is moving away to make room for it. Thus, an expansion setting the material in motion starts at a free surface and moves inward as a wave; the velocity of such a wave at any point may be shown (2) equal to the square root of the local value of the slope $dP/d\rho$ along the adiabat there. (NOTE: Slopes measured directly from Figure 1 are $d(\ln P)/d(\ln \rho) = (dP/d\rho) (\rho/P)$.) If the material is heated uniformly, one adiabat will characterize the entire volume of material; if not, the heat energy deposited at each point will determine the adiabat descriptive of that point. For simplicity, assume in our imaginary example that the material is semi-infinite with one free surface and that heating is uniform throughout at 2000 cal/g. At some chosen interior point, a high pressure will exist until sufficient time has elapsed for an expansion front to reach it traveling from the free surface inward at a speed appropriate to the adiabat (see Figure 1). The pressure and density at this point will then fall, following the steep course of the adiabat until the liquidus line bounding the mixed liquid vapor region is reached. Here, there is a discontinuity in the slope of the adiabat. The expansion wave velocity is decreased by an order of magnitude, reflecting the highly compressible nature of the foam formed as liquid turns to vapor.

Figure 2 shows this front and the succeeding fan and Figure 3 illustrates the effect of the delay between passage of the liquid expansion and arrival of the mixed phase expansion in maintaining the pressure or density. The large change in density accompanying the second fan results in a large decrease in the electrical conductivity of the material. This is responsible in exploding wires for current "dwells" (4) during which field strengths from 5-10 kV produce insignificant current flow.

3. EXPERIMENT

In the experiments to be described, the semi-infinite sample is approximated by surrounding the edges and back surface of a thin, wide sample ribbon with a dense support material. The front surface is irradiated by an intense electron beam and, being unsupported, is free to expand. To detect the inward running expansion wave, a nominally constant current is directed lengthwise through the sample; as the advancing wave consumes the unexpanded, conducting material, the sample resistance rises, thus developing a proportional voltage increase. Copper and aluminum samples, as used in our experiments, are shown in Figure 4. All samples were 0.51 cm wide; thicknesses ranged from 0.005 to 0.015 cm. Figure 5 shows a cross-section of the construction of these assemblies. The substrate is a composite consisting of a 0.32 cm thick copper base (1 cm wide by 6 cm long) to which a 0.005 cm layer of fused silica was attached with the thinnest practicable (~ 0.0025 cm) layer of epoxy cement. The thickness of the base was limited by the penetration time of the beam-guiding magnetic field to avoid distortion of that field near the sample. The sample, and also, on either side of it, fused silica retaining walls, were attached to the electrically insulating, fused silica portion of the substrate. On the back side of the base, a thin copper ground strap and lead flyer plates, $\sim 0.3 \times 0.7 \times 1.6$ cm, were attached. The ground strap connected the signal cable shields to a reference connection at the center of the sample--which was also, approximately, at the center of symmetry of the beam. The flyer plates were designed to break loose and to carry away momentum transmitted by compression waves traveling through the substrate and the ground strap. This prevented voltage inducing motion of the ground strap in the beam guiding field and diminished reflections of stress waves from the back surface of the base.

The exciter current of ~ 33 A was produced by a 1 ms RC discharge and during the ~ 10 μ s experiment time was constant within about 15 percent. Its connection to the ends of the sample is diagrammed in Figure 6. The sample's grounded center tap was necessary to satisfy a requirement for symmetry imposed by the nature of the electron beam; it effectively created two samples. Leads to carry the voltage signals produced by the vaporizing ribbon were attached near each end, distinctly separated from the current connections to avoid any contact voltages. These leads' coaxial shields, though not shown, were connected via the ground straps noted earlier to the center tap. A low pass (<10 MHz) filter in each lead removed the large inductive pulses accompanying turn-on and turn-off of the beam current. One of the dual beam oscilloscopes, triggered internally, recorded signals from both sample halves using a high sweep rate; the other, initiated

by a trigger from the pulser, recorded both signals at a slow sweep rate. This arrangement reflects difficulty experienced in achieving proper triggering.

The pulser used was the Physics International SNARK (5). Its beam guiding solenoid was pulsed prior to discharge of the two parallel-plate, water insulated lines that power the beam. As the beam passed through the solenoid's converging, longitudinal field, it was first constricted and intensified and then expanded and weakened (as indicated in Figure 6). Thus, by a change of position, the sample could be subjected to a variety of intensities (or fluences) (6); the range chosen, expressed in terms of fluence absorbed in the sample, was from 44 to 111 cal/cm². The beam energies varied from 689 to 843 keV. As accompanying effects, the mean angle of incidence of the electrons varied between 30° and 40° and the beam diameter, between 4.4 and 5.6 cm.

Figure 7 shows rear and front views of the sample holder. Below the graphite shield ring attached in the front (right hand) view, the diametrically oriented sample with its center tap screw in place may be seen. On either side of the sample, 1 cm circles--the tops of cylindrical, graphite calorimeters--are exposed. The back (left hand) view of the holder shows two 0.6 cm thick stainless steel plates between which the sample substrate was clamped. Shallow, circular recesses machined in each plate accommodated the ends of an array of thermoluminescent detectors (TLD), one of which is shown (below the rear-view) in its acrylic plastic positioner. A collimating aperture in the center of each recess was drilled through to the front surface of the plate. A 0.0025 cm thick tantalum foil covered the entire irradiated front surface, except for the areas of the calorimeters and of the sample assembly, and converted the electron flux to x-rays. After passing through a 0.15 thick graphite filter, which stopped all electrons, the x-rays entered the collimating apertures and then excited the TLD's. Later "reading" of the TLD's provided a measure of beam uniformity which, from shot to shot varied between $\pm 5\%$ and $\pm 20\%$ (6).

Figure 8 shows representative sample-response curves. The upper pair of records represents slow and fast sweep traces for a nominally low energy exposure. The noise level exhibited is the lowest experienced. The slow-sweep oscilloscope was pretriggered and the first burst of noise observed coincided with the triggering of the pulse line charging unit; the second burst coincided with the beam turn-on. The traces marked "Sample 1" and "Sample 2" represent responses of the two halves of the center tapped sample. The lower pair of records was representative of the worst of those considered

BURDEN

readable. The difficulty in determining amplitudes or initiation times accurately is evident. The reading approach used was to average the peak excursions, either numerically or visually, depending on their severity. The auxiliary sensing current records were correlated in time so that a resistance history could be deduced from the voltage signals.

4. INTERPRETATION OF DATA

An expression relating oscillographically measured quantities, on the left side, to physical quantities on the right

$$R^{-2} dR/dt = (w/l) [c(e)/\rho(e)] \quad (1)$$

may be easily derived using the two relations

$$R = \rho l/w(x_f - x)$$

and

$$c = dx/dt$$

where R = measured resistance (ohms), t = measured time (s), c = velocity of expansion front (cm/s), ρ = resistivity (ohm-cm), e = specific energy (J/g), w = width of sample (cm), x_f = thickness of sample (cm), x = depth below irradiated surface (cm), and l = length of sample (cm). Inductive voltages developed across the sample have been considered negligible in this treatment in part because the sensing current is not changing rapidly and in part because the rate of change of inductance (7) related to the changing thickness of the conducting material has been rendered small by the use of relatively broad ribbons. The latter voltage, in its worst case, is on the order of 2% of the measured resistive voltage. The sensing current distribution in the samples after heating is predicted by transient skin effect theory (8) to have effectively reestablished equilibrium in about 75 ns for 0.0102 cm thick copper samples and 125 ns for the 0.0152 cm thick aluminum samples--the thickest of each type. These times are on the order of the 100 - 120 ns beam duration times and, at most, a small percentage of the 1-10 μ s running times of the vaporization waves.

It should be noted that the heating is not uniform with depth, thus pressure gradients exist which may allow internal expansions; the front surface expansion appears, however, to remain dominant in pressure relief.

BURDEN

A discrete discontinuity in conductivity at the vaporization front has been implied in the derivation, but it appears likely that there is a more diffuse region of change centered somewhere in the expansion fan and traveling at a velocity close to that of the expansion front. Further, the heating is not instantaneous, but takes a small fraction of the expansion's traverse time. Nor is the medium supporting the wave stationary: the deposition is such that the front surface of the sample, having expanded through the solid and liquid phases, may be moving with a velocity on the order of the vaporization wave velocity, and front and rear surfaces may be diverging with a velocity one tenth of that. We neglect this relative motion and assume that the expanding substrate maintains pressure at the sample's back surface.

Two approaches to interpretation of the data have been used and each makes use of data from exploding wire experiments. These data have been fitted to polynomials and express wave speed and resistivity as functions of specific energy. For copper

$$c = 4229. - 6.534e + 0.02524e^2$$

and

$$\rho = 2.047 \times 10^{-5} [1 + 4.769 \times 10^{-4} (e - 1090.)].$$

For aluminum

$$c = 11,353. - 6.370e + 0.000876e^2$$

and

$$\rho = 4.045 \times 10^{-5} [1 - 1.620 \times 10^{-4} (e - 3548.)].$$

For each experimental value of $R(t)$, a value of e may then be found. Corresponding values of c may be calculated and integrated over time to yield values of x .

Such data is plotted in Figures 9, 10, and 11 which also compare similar curves determined independently, by more conventional methods at the Physics International Co. (6) The agreement in results suggests the use of ribbon samples as deposition profile gages.

An alternate treatment of the data provided a determination of the velocity function $c = c(e)$, quasi-independently of the exploding wire experiment. It had to depend on the resistivity

BURDEN

function $\rho = \rho(e)$ determined by that experiment since this type of information could not be extracted from the present experiment. The specific energy came from among six normalized deposition profiles calculated using a condensed case history Monte Carlo electron transport code. Profiles were computed for the combinations of two mean angles of electron incidence and three beam voltages. The values plotted were obtained after three steps: interpolation between profiles to actual conditions of the test, application of a correction for the fused silica substrate, and representation of the result by a two-segment, polynomial fit. (The "dosimetry" curves of Figures 9-11 are plots of these normalized curves scaled by measured fluence.)

The experimental resistance data was converted to velocity by the use, again, of Equation 1. Now a resistivity corresponding to a trial value of energy was used to calculate c and x ($=\Sigma c\Delta t$). This trial value was then compared with the polynomial value $e(x)$ and adjusted by an iterative loop until internal consistency was achieved.

Figures 12, 13, and 14 show these results and, for comparison, plots of the exploding wire wavespeed versus energy polynomials.

Fluence measurements were made using thermoluminescent detectors (TLD) (whose precision was $\pm 3\%$) calibrated against the average of a large number of carbon calorimeters (whose accuracy was $\pm 10\%$). Normalized deposition profiles were calculated with a precision estimated at $\pm 10\%$. Specific energies were thus quoted with an accuracy estimated at $\pm 15\%$ by the Physics International Co. Measurements employing a stacked carbon foil depth-dose gage, when scaled to the areal density of copper, generally agreed with the calculated profile (6).

The nominal precision of the exploding wire data is about $\pm 8\%$. The portion of the data used in determining the resistivity function occupied only a few percent of full scale; thus, it is not surprising to find variations from shot to shot of $\pm 30\%$. This points to a need for improving this data.

Neither the triggering nor the electrical noise isolation was good in the data obtained. Consequently, in the pre-triggered, slow sweep shots it was difficult to see the experiment initiation time to within $0.1 \mu s$; there was even greater uncertainty in the internally triggered, fast sweep shots since it is not clear which event actually triggered the sweep. Finally, noise almost totally obscured the first $0.3 - 1 \mu s$ after initiation and made the accuracy of reading of the early amplitudes 10 to 30% at best.

5. ASSESSMENT

Those results presented in Figures 9-14 are selected from 15 copper and 2 aluminum exposures; the copper results are generally representative of the extremes in deposited energies. This experiment, depending as it does upon externally generated data for its interpretation, serves to relate the sources of this data if the physical process is properly described. It appears (assuming that the external data is correct) that the description is good at energies near the vaporization threshold for copper ($\sim 1400\text{J/g}$) and not so good at energies twice this value. This trend, with exceptions, was observed throughout the data. The vaporization thresholds for the low energy copper and the aluminum samples (Figure 12 and 14) are remarkably close to those found in static, atmospheric-pressure experiments (9).

In relating exploding wire and electron beam results, notice should be taken of their relative heating times. The nominal heating time for the electron beam ($\sim 100\text{ ns}$) is $1/5$ to $1/10$ that for the exploding wire, and the corresponding trajectories through the nonlinear state diagram are different. Thus, it is probable that a different value of the terminal velocity of the medium results from the expansion through the solid and liquid phases.

An interesting correlation occurred in three cases between pre-exposure observations of improper bonding of sample or side retaining walls and failures of the half of the sample in which it occurred. This is interpreted as evidence that the support material was performing the intended function and that, normally, expansion from the free surface did account for sample response.

A large potential source of error in results lies in the resistivity function, whose error bounds are large. In fact, a 20% increase in resistivity brings experimental deposition profiles into notably better alignment with the dosimetric values for copper. However, except for the data of Knoepfel and Luppi (10) which is on the order of 10% higher, no justification exists for revising it. More accurate determinations of this function are called for.

The integration of calculated velocities as reflected in the x-coordinates of Figures 9-11 provides an interesting and demanding self-check on the validity of the interpretation. Although considerable scatter is found between the calculated depths of penetration and the vaporization threshold depths of the transport code, in no case does the penetration of the wave exceed the thickness of the sample. This is despite the errors introduced into the higher,

early time velocities by electrical noise and may in part be attributed to the highly stable nature of the e versus x calculations which tends to correct for earlier velocity errors.

6. CONCLUSIONS

The program described has provided a means for experimental contact with the adiabats in the mixed-liquid-vapor regime of the equations of state for two metals. Because of the high temperatures and pressures necessary, these adiabats are not accessible by static processes.

The significant achievement of the series of tests is that a thermodynamically described phenomenon, which had served to explain the response of exploding wires, could be used to predict the unique response actually found in samples irradiated by a vaporizing electron beam. It is also significant that, when analyzed, this response could be inverted to describe to a reasonable, though not high, degree of accuracy the input conditions producing it.

The model used and the experimental techniques require refinement. Computer hydrocodes with realistic equations of state should provide a more precise model and an effort is being commenced in which one or more modifications to the equations of state available to the RIP hydrocode will be made. Options are (a) incorporation of the Goodwin equation of state (3), (b) revision of the presently incorrect tabular mixed-liquid-vapor sound speeds used in the RIP equation of state (1), and (c) modification of the GRAY (11) equation of state to describe the mixed-liquid-vapor region using a physically consistent approach developed by Kahl (12).

Also, further more accurate work in determining the resistivity function and in obtaining cleaner experimental signal traces is planned.

BURDEN

REFERENCES

1. R. H. Fisher and H. E. Read, "RIP, A One-Dimensional Material Response Code," SSS-R-72-1324, Sep 72 (Systems, Science, and Software, La Jolla, California 92037).
2. F. D. Bennett, G. D. Kahl, and E. H. Wedemeyer, "Resistance Changes Caused by Vaporization Waves in Exploding Wires," in Exploding Wires, Vol. 3, W. G. Chace & H. K. Moore, Ed., (New York, Plenum Press) 1964.
3. L. K. Goodwin, L. A. Johnson, and R. S. Wright, "An Equation of State for Metals," DASA-2286, Apr 69 (Aeronutronic Div. of Philco Ford Corp., Newport Beach, California 92660) NTIS Order No. AD858302L.
4. R. J. Reithel, J. H. Blackburn, G. E. Seay, and S. Skolnick, "The Current Pause in an Exploding Wire," in Exploding Wires (Vol. 1.) W. G. Chace and H. K. Moore, Ed., (New York, Plenum Press, Inc.) 1959.
5. S. Shope, P. Spence, and C. Stallings, "SNARK Electron Beam Development for Materials and Structures Irradiation," PIFR-301, June 1972 (Physics International Co., San Leandro, California 94577).
6. P. Spence, "Electron Beam Test Support for Mixed Phase Expansion Velocity Measurement," PIFR-362, Sep 72 (Physics International Co., San Leandro, California 94577).
7. Page 5-28, American Institute of Physics Handbook, Dwight E. Gray, Coord. Ed., (New York, McGraw-Hill) 1957.
8. F. D. Bennett and J. W. Marvin, "Current Measurement and Transient Skin Effects in Exploding Wire Circuits," Rev. Sci. Inst. 33, 1218 (1962).
9. See Ref. 7, pages 4-42 and 4-130.
10. H. Knoepfel and R. Luppi, "The Electrical Conductivity of Metals at Very High Temperatures," in Exploding Wires, Vol 4, W. G. Chace and H. K. Moore, Ed., (New York, Plenum Press) 1968.
11. E. B. Royce, "Gray, A three-Phase Equation of State for Metals," UCRL-51121, September 1971.
12. G. D. Kahl, "Generalization of the Maxwell Criterion for Van Der Waals Equation," Phys Rev 155 (2), 78 (1967).

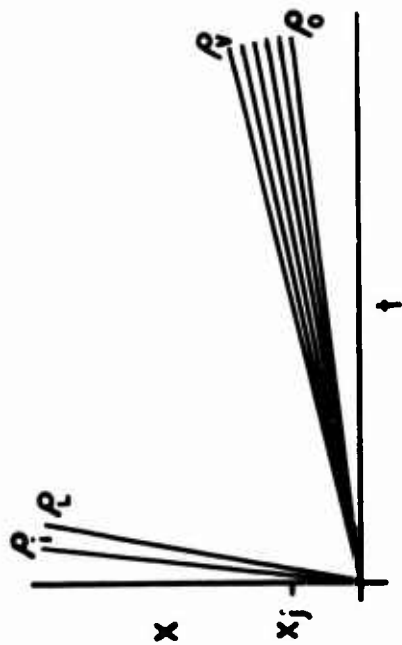


FIGURE 2. Expansion Fans on Depth (x) Versus Time (t) Plot Representative of Liquid (ρ_i to ρ_L) and Mixed-Liquid-Vapor (ρ_v to ρ_o) Expansions.

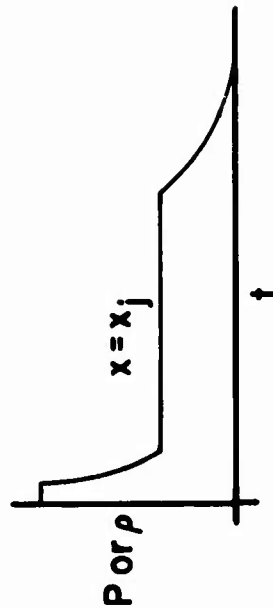


FIGURE 3. Qualitative Representation of Pressure (P) or Density (ρ) versus Time at Point x_j of Figure 2. (Vertical Scales Highly Nonlinear).

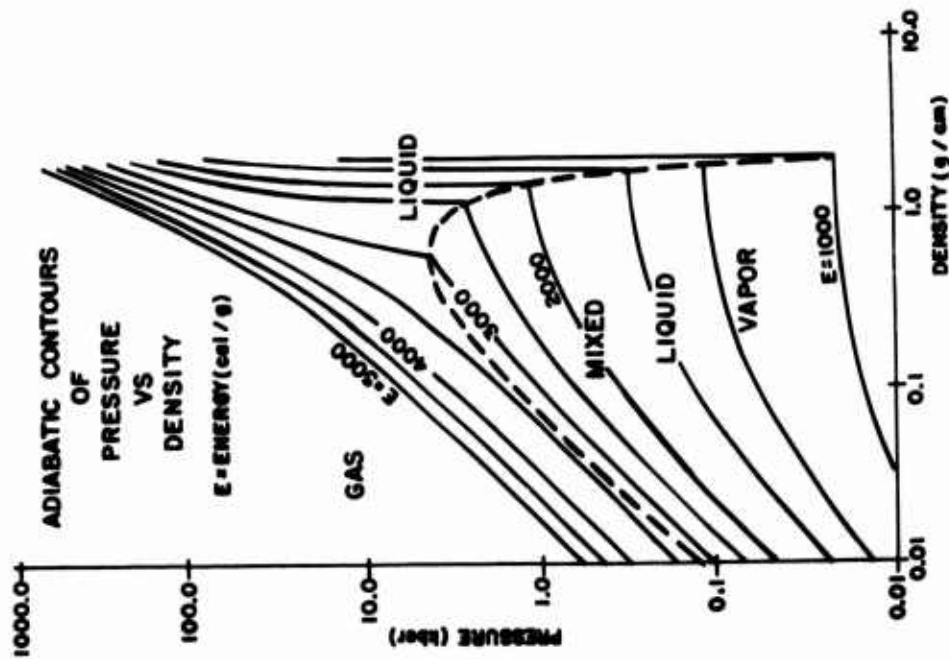


FIGURE 1. Pressure-Density (P - ρ) Diagram for Aluminum (3) Showing Constant-Energy (Adiabatic) Contours.



FIGURE 4. Copper and Aluminum Ribbon Samples Mounted on Copper Base Blocks.

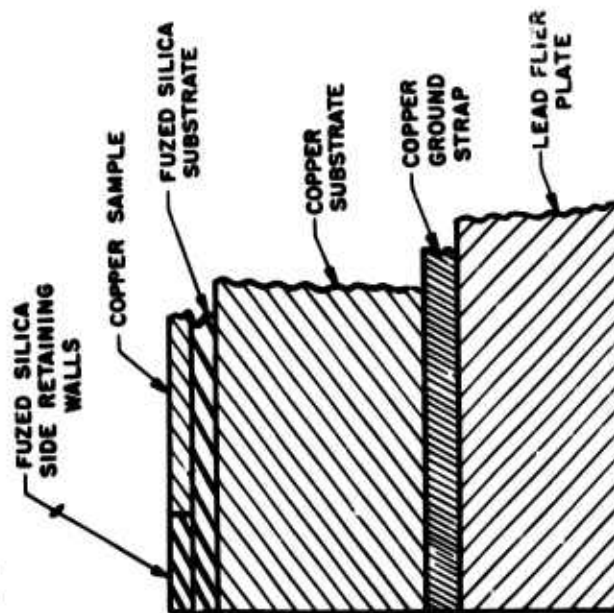


FIGURE 5. Diagram of Sample Holder Assembly. (Not to Scale.)

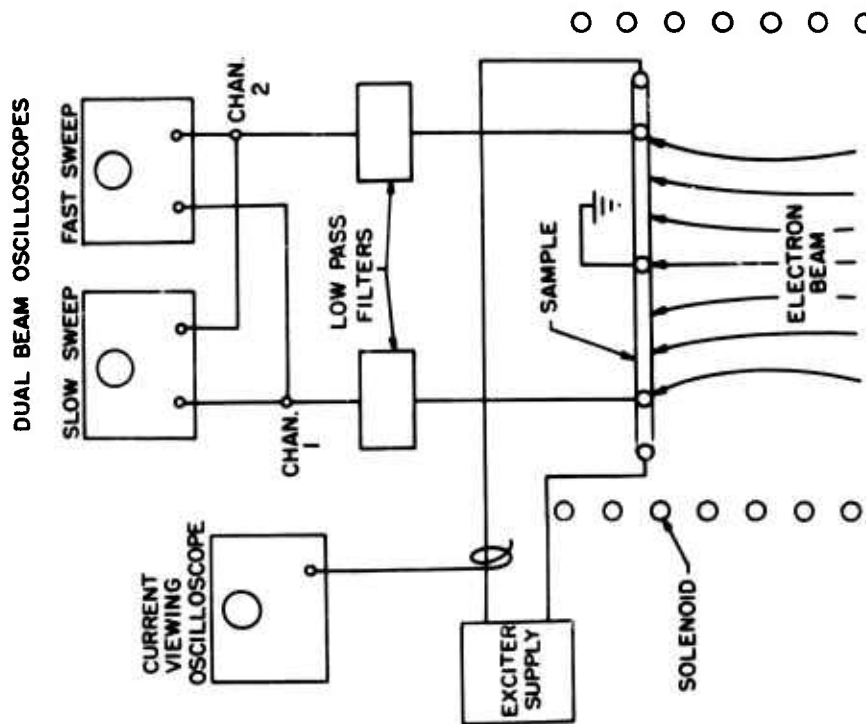


FIGURE 6. Basic Elements of Experimental Arrangement.



FIGURE 7. (Right) Sample Holder Completely Assembled. (Upper Left) Rear Face of Disassembled Holder Showing TLD Recesses. (Lower Left) Typical TLD Array.

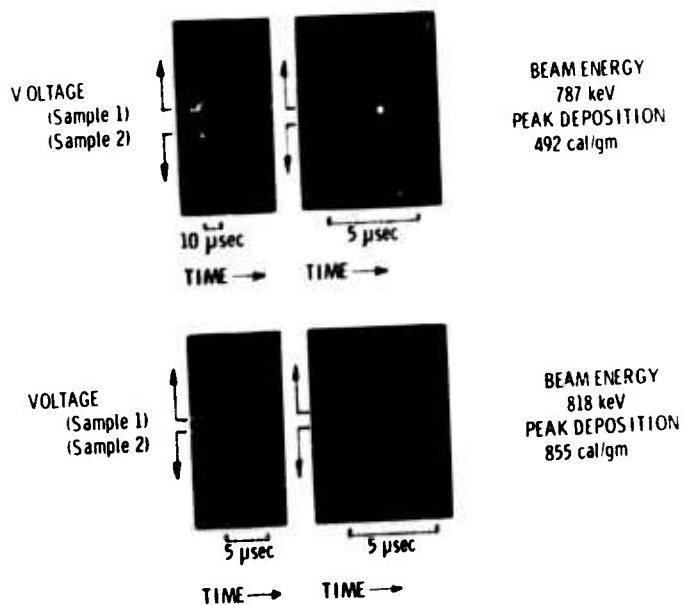


FIGURE 8. Oscillographic Data Trace Sets Showing Voltage Response of Two 0.0056 cm Thick, Center-Tapped Ribbons at Slow and Fast Sweeps.

BURDEN

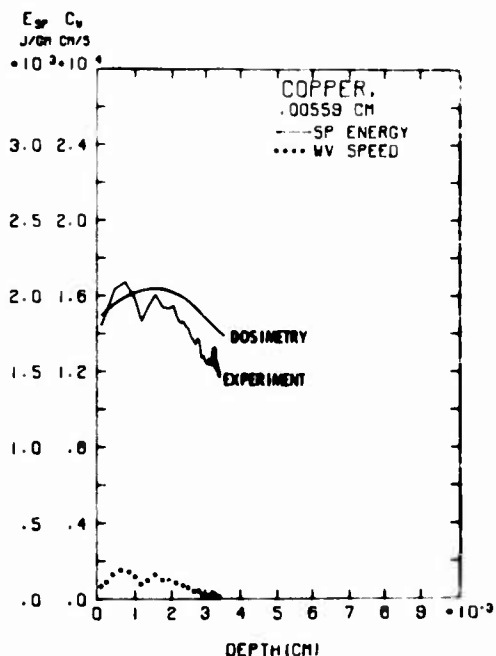


FIGURE 9. Specific Energy (E_{sp}) and Wave Speed (C_w) versus Depth in Copper at 48 cal/cm² Absorbed.

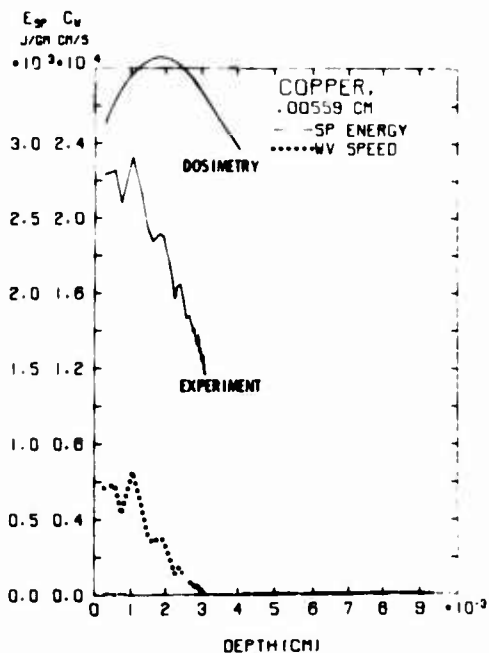


FIGURE 10. Specific Energy (E_{sp}) and Wave Speed (C_w) versus Depth in Copper at 90 cal/cm² Absorbed.

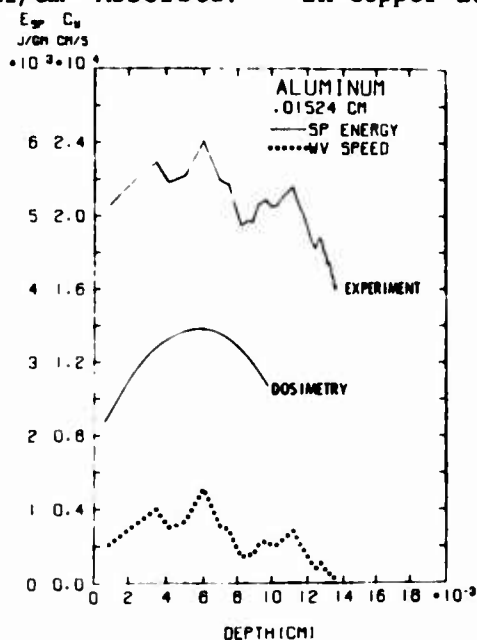


FIGURE 11. Specific Energy (E_{sp}) and Wave Speed (C_w) versus Depth in Aluminum at 91 cal/cm² Absorbed.

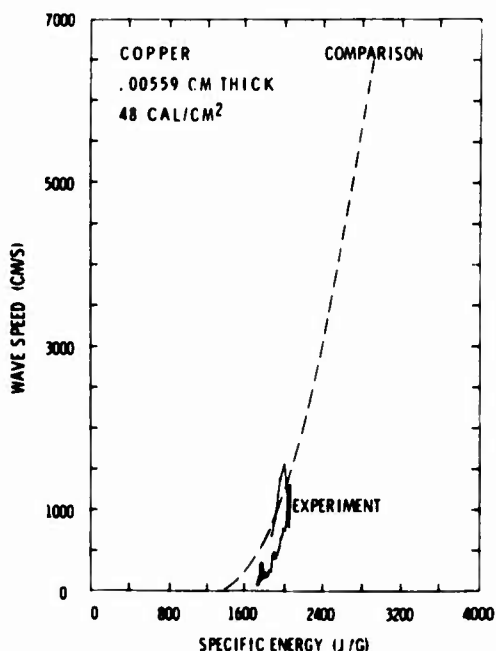


FIGURE 12. Measured Wave Speeds versus Specific Energy Compared with Exploding Wire Data Fit.

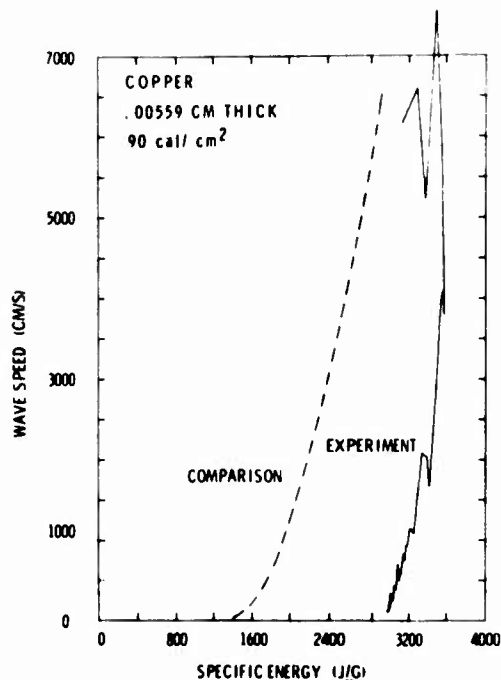


FIGURE 13. Measured Wave Speeds versus Specific Energy Compared with Exploding Wire Data Fit.

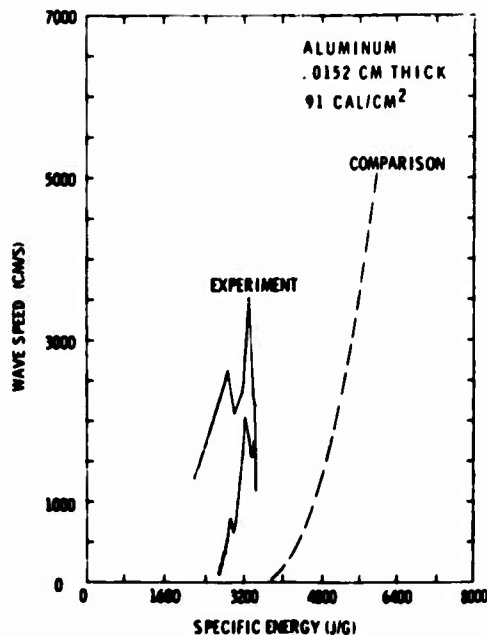


FIGURE 14. Measured Wave Speeds versus Specific Energy Compared with Exploding Wire Data Fit.

DYNAMIC EFFECTS IN THERMAL BLOOMING IN HIGH POWER
LASER PROPAGATION

RUDOLF G. BUSER, PhD
ROBERT S. ROHDE, PhD
COMBAT SURVEILLANCE & TARGET ACQUISITION LABORATORY
US ARMY ELECTRONICS COMMAND, FORT MONMOUTH, N.J.

1. INTRODUCTION

1.1 Problem Definition

For military applications, correct prediction of atmospheric effects upon laser propagation is important. For pulse propagation no unified theory exists and limited (collimated beam), or no (focused beam) experimental verification is available. An objective must be to establish, verify, and extend existing theory for pulse propagation, such that optimized propagation conditions may be determined.

To provide guidance for a comprehensive theory, including collimated and focused Gaussian beam geometries, as well as others (for example, the unstable resonator configuration) an extensive experimental program has been undertaken, and phase 1, which deals with pulse durations longer than the acoustic transit time, has been completed. The results obtained and discussed in the following indicate that the performance of collimated beams is adequately described by existing theory if thermal conduction is included; however, considerable departures are observed for focused beams.

1.2 Background

If a laser beam of sufficient intensity is turned on in an absorbing medium, an initially transient thermal lens is produced and concomitant changes in the energy distribution of the beam in the target plane are observed. Details of this process strongly depend on beam geometry, energy transfer, and transport phenomena within the medium. Theoretically, the development of transient thermal blooming has been investigated for specific conditions by several authors⁽¹⁻⁵⁾ and involves a simultaneous solution of the hydrodynamic equations and the wave equation by (usually) numerical techniques. No generalized approach presently exists which extends from initial laser switch-on to steady state conditions and includes forced and induced convection, translation and slewing, and turbulence induced phenomena as required in a variety of propagation scenarios. Rather, one is forced to consider segments of the total problem separately and to introduce various simplifying assumptions in the development of a manageable theoretical model. The question arises as to which theoretical treatment is used for comparison: Preference has been given to the development presented in References 1 through 4 because the subject matter is derived most transparently and consistently and lends itself easily for adoption to experimentally required compromises. As pointed out, from an experimental point of view, few results are available⁽⁵⁻⁷⁾ and more data in support of existing theory and its further development is clearly needed.⁽⁸⁾

2. DISCUSSION OF EXISTING THEORIES

2.1 Collimated Beams

In a first order treatment of transient collimated beams⁽²⁾ the following assumptions and approximations are made: (a) the medium is initially quiescent, homogeneous, and isotropic, i.e., wind, convective currents, and density fluctuations are ignored. (b) Induced convection due to heating of the medium is disregarded, and the theory is only applicable to times short compared with the time for convection to set in.⁽⁹⁾ (c) Properties of the medium can be expressed in terms of the linearized hydrodynamic equations. (d) The changes in beam intensity due to absorption will be ignored. (e) Geometrical optics (no paraxial approximation) is used to describe the beam and diffraction effects are ignored. (f) A heat source which acts instantaneously at all points in space simulates the heating effect of the laser beam the instant it is switched on. For Gaussian profiles and negligible thermal conductivity a scalable numerical solution is available which may serve as an initial experimental guide. Specifically, the normalized intensity distribution is inversely proportional to the beam area (each ray contributing to the

total intensity) as given by

$$I_{N,C}(x, \xi) = \sum_{x_0} \frac{f(x_0)}{\left| \left(\frac{dx}{dx_0} \right)_{\xi} \right|} \cdot \frac{x_0}{x} \quad (1)$$

with $x_0 = r/a$, $x = r/a$, a the (1/e beam radius), and the sum over all x_0 which yield the same value of ξ for a given x . For a pair (x_0, x) , ξ is computed from

$$\xi = -2 \frac{\sqrt{f(x_0) - f(x)}}{f'(x_0)} - 2 \int_{x_0}^x dx' \frac{f''(x') \sqrt{f(x_0) - f(x')}}{[f'(x')]^2} \quad (2)$$

and

$$\left(\frac{dx}{dx_0} \right)_{\xi} = + \frac{f'(x_0)}{f'(x)} - \frac{2f'(x_0)f''(x)}{[f'(x)]^3} [f(x_0) - f(x)] \\ + 2 \sqrt{f(x_0) - f(x)} f'(x_0) \int_{x_0}^x dx' \sqrt{f(x_0) - f(x')} \quad (3)$$

$$+ \frac{f'''(x') f'(x') - 3 [f''(x')]^2}{[f'(x')]^4}$$

For the Gaussian beam,

$$f(x) = e^{-x^2} \quad (4)$$

The quantity ξ is related to the physical parameters, neglecting viscosity and thermal conduction, by

$$\xi = \left[\frac{1}{3} \frac{(n_0^2 + 1)(n_0^2 + 2)}{n_0^2} \cdot \frac{3\beta c^2}{2c_p} \cdot \frac{t}{t_c} \right]^{\frac{1}{2}} \frac{z}{a} \quad (5)$$

with

$$t_c = \frac{3\pi a^2 n_0 c^2}{2\alpha p} \quad (6)$$

which is the characteristic time required for the laser to cause a density change in a perfect monatomic gas to be equal to $\pm \rho_0$. In the above equations n_0 is the initial index of refraction, z the path length (cell length for the experiments), ρ_0 is the initial fluid density, β its volume expansivity, c_p the specific heat at constant pressure, α the absorption coefficient, c the speed of sound, P the total power, and t the time.

Within the range of applicability of equation (2) ($\xi < 3.0$), there is only one value of x_0 yielding a ξ for a given x . Equations (1) - (4) have been solved numerically for two cases $x = 0, 1$ and the results are displayed in Figure 1.

The assumption of negligible thermal conduction is applicable only for times $t < a^2/D_f$ (D_f diffusivity). If thermal conductivity is included in the model, the numerical solutions become nonscalable, and for each experimental case a complete numerical analysis has to be made. (2) For a parabolic beam profile

$$f(x) = \begin{cases} 2(1 - x^2) & x \leq 1 \\ 0 & x > 1 \end{cases} \quad (7)$$

and the on-axis intensity (neglecting thermal conduction) has an analytic solution and is given by

$$I_N = \frac{I(0, z, t)}{I_0} = \frac{2}{\cosh^2 \sqrt{2\xi}} \quad (8)$$

For the annular beam profile

$$f(x) = \begin{cases} 0 & r \leq b \\ e^{-x^2} & r > b; x = r/a \end{cases} \quad (9)$$

where b is the radius of the blocking aperture, fundamental computer problems are encountered if the outlined scheme is followed, and further work is required. It is interesting to note that similar to the previously introduced distortion parameter for cw beams steady state conditions, (10, 11) a pulse distortion parameter N_p may be defined following the expression

$$I_N = \exp(-N_p) \quad (10)$$

where N_p for times long as compared to the acoustic transit time is defined by

$$N_p(t) = \frac{2}{\pi} \frac{(-dn/dT) \alpha z^2 P t}{n_o n_o c_p a^4} \quad (11)$$

The usefulness of this expression is contained in the fact that propagation engineering nomograms may be constructed which permit fast estimates of propagation effects in a given scenario. It also follows, that for the collimated beam

$$\text{or } I_N \sim \alpha P t z^2. \quad (12)$$

2.2 Focused Beams

For the focused beam the same initial assumptions made for the collimated beam are used with the exception that the effects of diffraction are included. For the Gaussian profile the intensity is given by

$$I_o = (P/\pi a^2 D) \exp \left[-r^2/a^2 D \right] \quad (13)$$

which includes the effects of focussing and diffraction via

$$\left[D(z) \right]^{\frac{1}{2}} = 1 - z/f + \frac{z}{(ca^2)}, \quad z/f \leq 1 \quad (14)$$

For the long time case the geometric optics calculation (the wave optics calculation leads to rather unmanageable integrals) yields

$$I_{N,F} = I/I_o(r,z,t) = 1 - \beta_{\infty}(z) \left[1 - \frac{2r^2}{a^2 D(z)} \right] \times \exp \left[-\frac{r^2}{a^2 D(z)} \right] \quad (15)$$

with β_{∞} a measure of the energy deposited in the medium and, therefore, of the degree of the perturbation with

$$\beta_{\infty}(z) < \delta_{\infty}(z) \frac{P t}{E_f}$$

δ_{∞} being a numerical parameter given by

$$\delta_{\infty}(z) = \frac{4}{k^2 a^4} \int_0^z \frac{dz'}{D(z')} \int_0^{z'} dz'' \frac{\exp(-\alpha z'')}{D(z'')} \quad (16)$$

and E_f a constant containing the hydrodynamic parameters

$$E_f = \pi c^2 / (3/2 N(\gamma - 1) \alpha k^2) \quad (17)$$

where f is the focal length, N the refractivity, γ the ratio of specific heats, and $k = 2\pi/\lambda$ with λ the wavelength. The equation for $\delta(z)$ must be solved numerically unless $\alpha \ll f^{-1}$ in which case $\exp(-\alpha z) \approx 1$ for $z < f$ and standard tables may be used. The following points appear pertinent:

(a) The method used in the solution is not self-consistent and implies straight rays converging to a point beyond the focal point in such a manner that the focal spot size is that of the wave optics beam (geometric optics). In the focal plane the value for δ_∞ will differ considerably in the long time limit for geometric and wave optics solutions, and the exact value has to be determined experimentally. Outside the focal plane both approximations produce equivalent results.

(b) A comparison of the focused with the collimated case, by letting $f \rightarrow \infty$, yields only comparable results for $I/I_0 \approx 1$, but deviates considerably for $I/I_0 \ll 1$. This problem is connected with the way diffraction has been introduced into the model.⁽¹²⁾

(c) Only for small values of β_∞ , say $\beta_\infty \leq 0.3$ quantitative agreement is expected. For larger values the used model of energy deposition by an undistorted beam is invalid.

(d) No theory exists for the focused beam including thermal conduction.

3. EXPERIMENTAL PROCEDURE, RESULTS, AND DISCUSSION

3.1 Experimental Setup

The experimental arrangement used for the pulsed propagation studies is shown in Figure 2. The laser source is a Coherent Radiation CO₂ Model 41 with internal power meter operating at 10.6 μ in the 10-100 W/cm² range in single mode (selectively P18, P20, or P22) close to diffraction limited. The collimated output beam is electro-mechanically shuttered (approximately 7 ms opening time) and expanded by an off-axis mirror telescope. The collimated or focused beam passes through a 2.15 m absorption cell. A translating mirror selects the profile segment to be scanned by a high speed mirror (8000 rpm) across an apertured (0.1 mm diameter hole) liquid

nitrogen cooled mercury-cadmium-telluride photoconductive detector. The received signal is amplified by a 1A7A plug-in unit in a 5458 Tektronix oscilloscope and is recorded on movie film (3600 in/min). The detector itself is located 2.85 m away from the entrance window of the cell. In all comparisons with the theory the length of the cell was used as the relevant interaction length. An original representative display of the data showing the time development of the central on-axis (0, Y) profile of the beam is shown in Figure 3. Repetition of the experiment at (X_1, Y) permits construction of the complete beam profile as a function of time. For the study of annular beams, mirror M2 was replaced by another mirror with a precision central aperture.

An appropriate choice of the experimental parameters is highly significant if a valid comparison with theory is to be accomplished. Estimates have shown that with the selected power densities for laboratory studies a convenient separation of long term behavior, induced convection, and steady state conditions is possible. It should be pointed out that in view of the 4th power dependence on beam radius for the focused case (equation (16)), a high degree of experimental care in the performance of the experiments is necessary if a quantitative comparison with theory is to be attempted.

3.2 Experimental Results and Discussion

3.2.1 Collimated Beams

Figures 4 and 5 show a comparison of the transient behavior of on-axis (0,0) and off-axis (0,a) intensities with theory for the Gaussian beam, the parameter varied being the laser power. All other values were kept constant:

$$\begin{aligned} \beta &= 0.0037 \text{ } ^\circ\text{C}^{-1}, \alpha = 2 \times 10^{-1} \text{ m}^{-1}, a = 4.52 \times 10^{-3} \text{ m}, \\ c &= 264.196 \text{ m/sec}, \rho_0 = 1.859 \text{ kg/m}^3, c_p = 0.842 \times 10^3 \\ &\text{joule k}^{-1} \text{ } ^\circ\text{C}^{-1}, n_0 - 1 = 4.58 \times 10^{-4}. \end{aligned}$$

Since the beam is slightly divergent the value used for the beam radius is that at the mid-point of the cell.

Agreement between theory (as presented by Equations (1-6)), and experiment is acceptable for early times. For longer times thermal conduction has to be included. The data show a better fit to theory for the on-axis intensity than for the off-axis values. This should

be expected since the Gaussian profile has a steep slope at $r = a$, and small deviations from an ideal Gaussian tend to be more accentuated than in the case of the peak where the slope is flat.

An attempt was made to also compare the data with the theory for the parabolic profile. Using the e^{-1} radius, Equation (8) predicts an intensity fall off faster by about a factor of 2 than the experimental data, as indicated in Figure 5. At first sight this appears to be in disagreement with previous comparisons of experimental data with theory in Reference 3 where acceptable agreement with the parabolic profile was obtained. However, careful analysis shows that with correct values for the absorption coefficient and beam radius the agreement is at best marginal for the parabolic but excellent with the Gaussian profile as it should be.

The decay of collimated annular beams proceeds slower as compared to Gaussian beams of equal intensity (Figure 6). In contrast to the collimated beam the decrease of the intensity initially appears linear, and is related to the diffraction forming effected by the annular aperture.

3.2.2 Focused Beams

Figures 7 and 8 show the representative results for transient on and off axis intensity decay ($X = 0, Y = 0; X = 0, Y = \pm a$) for focused beams for different absorption coefficients with constant input beam power (Gaussian beam). The gases used were mixtures of CO_2 and N_2 . The only other parameter used in the equations different from the previous set is the input radius r_i (at cell entrance) and the radius at the detector r_d (for vacuum propagation). For the experimental data presented these values were measured to be $r_i = 0.916$ cm and $r_d = 0.89$ mm, where the bar indicates an average value applying this definition; the measured $1/e$ I_0 area is set equal to $\pi \bar{r}^2$ where $a \equiv \bar{r}$. The theoretical curves are a best fit to the experimental data; the values used for r_i and r_d deviate from the experiment by approximately 10%. However, a 10% variation in radius translates into 40% variation in $1/I_0$.

The results obtained indicate that up to a certain $1/I_0$ value, the behavior is strictly linear as expected, and good agreement with the theoretical model is obtained. However, as the intensity continues to decay the linear behavior changes into an approximately exponential decay. The deviation from linearity becomes pronounced for all Gaussian on-axis measurements at $1/I_0 \approx 0.7$, again in agreement with theory and independent of the time when $1/I_0 = 0.7$ is reached. The question arises as to whether this agreement is

Incidental and the observed effect is perhaps due to processes connected with the initiation of induced convection. However, evaluation of measurements at various points ($X = 0, Y = 0$; $X = 0, Y = +a$; $X = 0, Y = +1.5a$; etc.) in Figure 9 shows that, while the deviation from the linear decay takes place at increasingly later times ($\alpha P t = \text{constant}$) as profiles with increasing Y are observed, the beam symmetry is maintained. This necessarily excludes induced convection as the culprit for the change in decay rate. Qualitatively it might be argued that as the focused beam blooms, it changes towards a collimated beam, the decay time behavior of which is exponential as discussed previously. Theoretical treatment of this problem presently does not exist.

As pointed out the experimental method permits the measurement of the development of the complete profile from opening of the shutter until steady state is reached. In Figure 10 the e^{-1} contour for this complete development is shown. Though the present discussion is concerned with times typically in the order of 100 ms, it should be kept in mind that the development towards steady state is a continuous, extremely complex process.

In Figure 11 the decay of the on-axis intensity of an annular beam (the ratio of annular hole diameter to entrance beam diameter $\epsilon = 0.63$) of same input power and for the same propagation conditions as shown in Figure 9 is displayed. The initially linear behavior of the intensity decay is again maintained but proceeds with a greatly reduced rate. At later times the intensity becomes also exponential but again falls off considerably more slowly than the Gaussian beam. Quantitative comparison with the theory requires extension of the present theory and work towards this end is underway. The results appear to indicate the the linearity of the intensity decay of focused beams is rather independent of the specific beam profile under consideration. In Figure 12 the e^{-1} contour's transient behavior for the complete transient period is presented. In rather dramatic contrast to the Gaussian beam, the e^{-1} contour remains approximately constant in size and induced convection processes are greatly reduced. Implications of this effect upon steady state conditions have been presented elsewhere. (7)

While laboratory simulation of transient beam spread indicates satisfactory agreement with existing theory, one might question the degree to which the theoretical model is scalable. Therefore, experimental results previously obtained with a 100 kW, 10.6 μ cw laser focused at 80 m atmospheric propagation path are presented in Figure 13. Details of the experimental setup have been discussed in Reference 13.

BUSER & ROHDE

Again the theory predicts the linear beam decay well within experimental accuracy in spite of the fact that the beam quality was limited and the energy distribution within the beam irregular.

4. CONCLUSIONS

Existing theory predicts well symmetric intensity decrease for Gaussian and annular stationary collimated beams. Predictions for the performance of focused beams is also satisfactory for $I/I_0 > 0.7$, but deviates substantially for $I/I_0 < 0.7$. The point of deviation is dependent upon α Pt, the energy deposited in the medium.

5. ACKNOWLEDGEMENT

Assistance of H. Gauch and I. Gumeiner in the data evaluation process and the assistance of J. Herder in the numerical calculations is appreciated.

6. REFERENCES AND FOOTNOTES

1. P.B. Ulrich, J. Opt. Soc. Am. 62, 1383A(1972).
2. J.N. Hayes, Appl. Opt. 11, 455 (1972).
3. P. B. Ulrich, J.N. Hayes, and A. H. Aitken, J. Opt. Soc. Am. 62, 728A (1972).
4. A.H. Aitken, J.N. Hayes, and P.B. Ulrich, Appl. Opt. 12, 193 (1973).
5. J. R. Kenemuth, C.B. Hogge, and P. Avizonis, Appl. Phys. Lett. 17, 220 (1970).
6. H. Kleiman and R.W. O'Neill, 'Appl. Phys. Lett. 23, 43 (1973).
7. R.G. Buser, R.S. Rohde, F.G. Gebhardt, and D.C. Smith, Conference on Laser Engineering and Applications, Washington, D.C., June 1973.
8. P.B. Ulrich, J. Opt. Soc. Am. 63, 897 (1973).
9. A lower limit on the thermal conduction time is given by

$$t_{\text{conv}} \geq [(4 \rho_0 c_p a \cdot \pi a^2)^{1/3} / (Bg\alpha P)]$$

where g is the acceleration due to gravity and the other parameters are as defined in the text.

10. F.G. Gebhardt and D.C. Smith, IEEE J. Quant. Elect., QE-7, 63 (1971).
11. R.G. Buser and R.S. Rohde, Appl. Opt. 12, 205 (1973).
12. P.B. Ulrich, Private communication.
13. R.G. Buser, R.S. Rohde, F.G. Gebhardt, and D.C. Smith, 5th DOD Laser Conference, Monterey, California, April 1972.

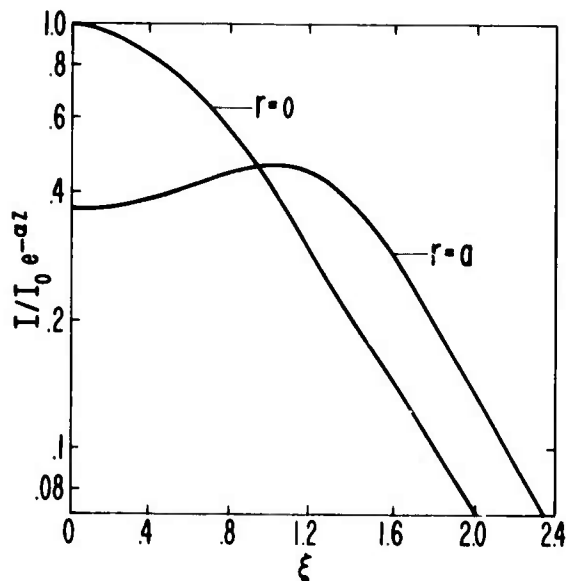


Fig. 1. Calculated normalized intensity for collimated gaussian beam as a function of ξ for two values of r ; thermal conduction not included.

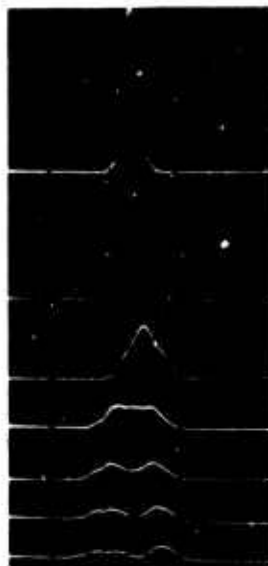


Fig. 3. Representative experimental data display of temporal evaluation of the $(x=0, y)$ profile; time between selected traces, 40 ms; individual trace, 20 μ s.

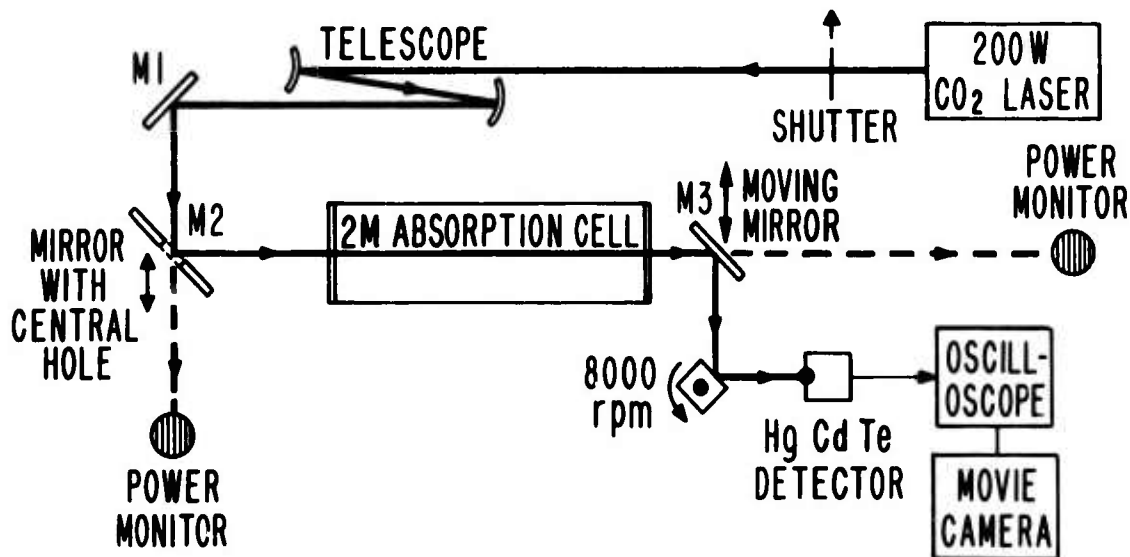


Fig. 2. Experimental test set-up.

Fig. 4. Normalized beam intensity N_c as a function of time; collimated beam; CO_2 ; $P = 18.5 \text{ W}$; experimental points $x(r=0)$, $\bullet(r=a)$; — gaussian theory; — — — gaussian theory including thermal conduction.

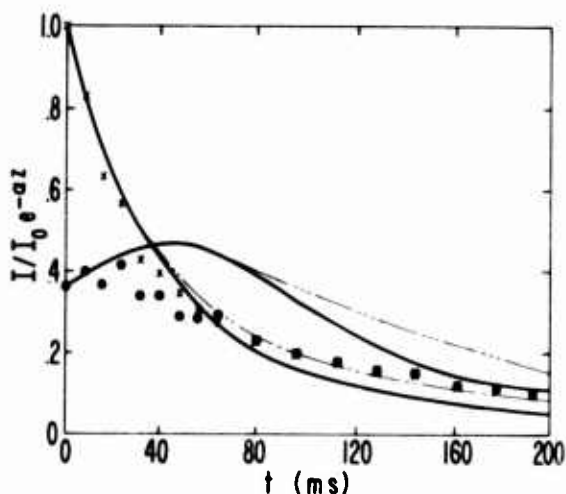
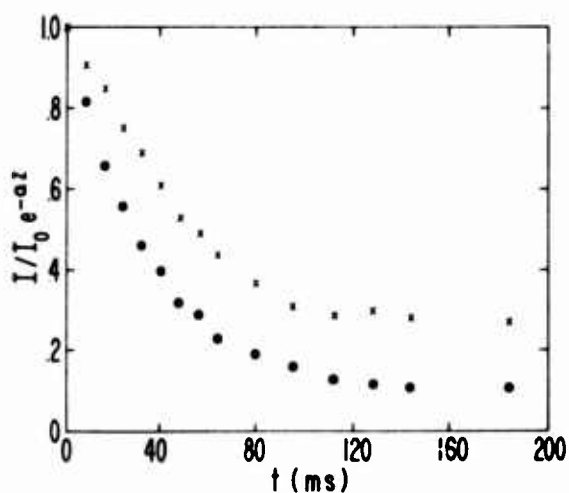
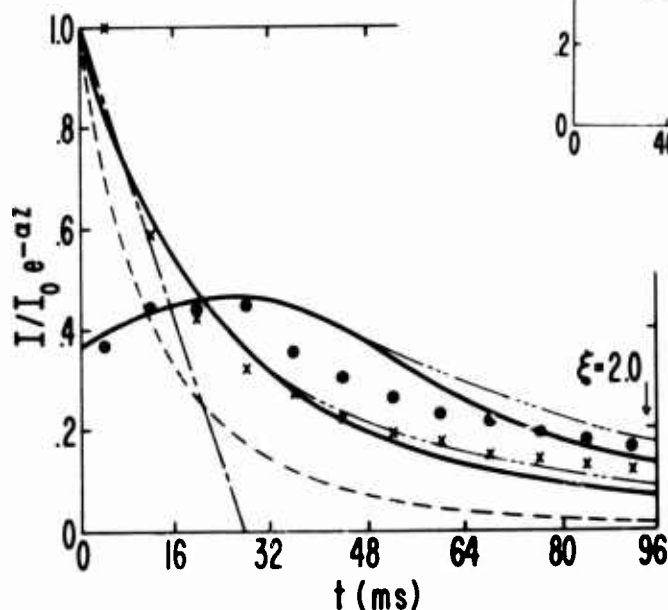
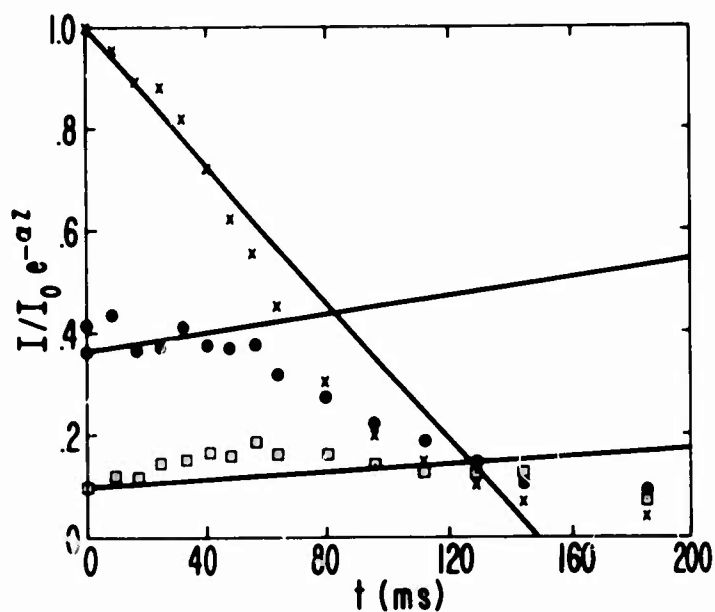
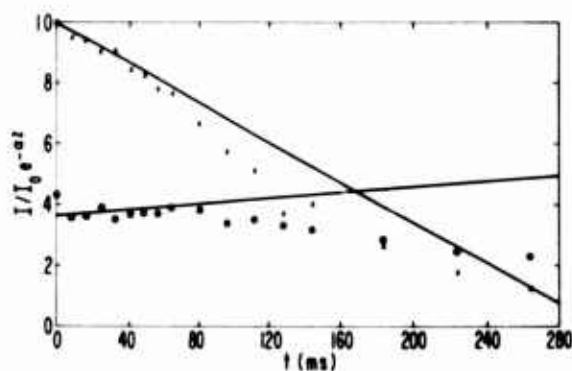
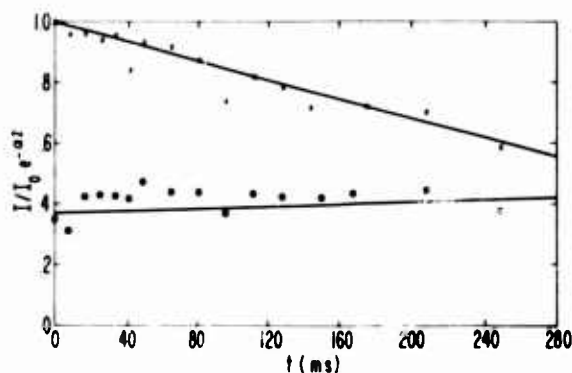


Fig. 5. Normalized beam intensity $N_c = f(t)$; collimated beam; CO_2 ; $P = 33.4 \text{ W}$; $x(r=0)$, $\bullet(r=a)$; — gaussian theory; — — — gaussian theory including thermal conduction; — — — parabolic theory; — — — focused theory with $f = \infty$.

Fig. 6. Normalized peak intensity $N_c = f(t)$; collimated annular beam; CO_2 ; $\epsilon = 0.6$; $x(P=18.5)$, $\bullet(P=36\text{W})$.



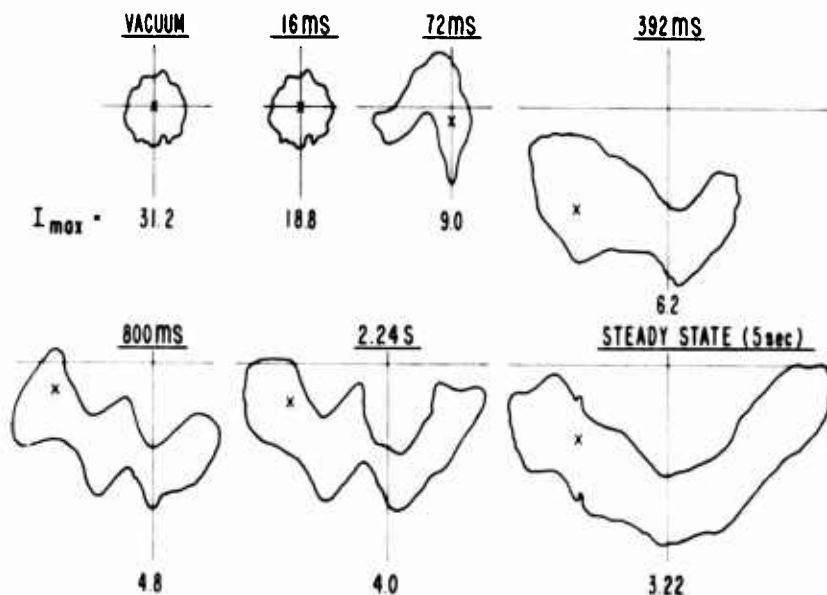


Fig. 10. Focused gaussian beam temporal evolution of $1/e I_0$ profile; CO_2 ; $P = 25.7 W$; I_{max} , peak height in relative units; x , position of peak.

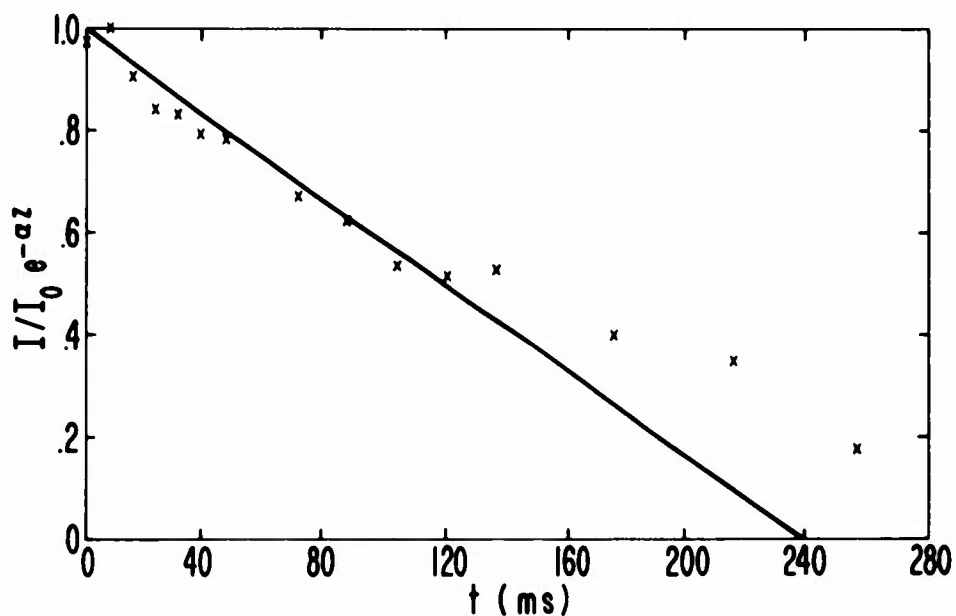


Fig. 11. Normalized beam intensity $N_f = f(t)$; focused annular beam; CO_2 ; $P = 25.7 W$; $e = 0.65$; — best fit to linear portion of data.

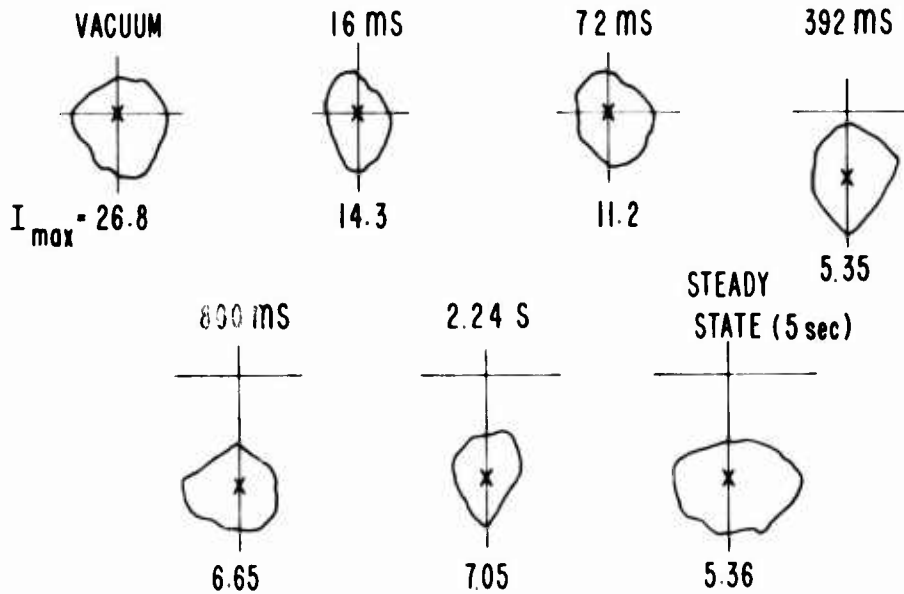


Fig. 12. Focused annular beam temporal evolution of $1/e$ I_0 profile; CO_2 ; $P = 25.7$ W; I_{\max} , peak height in relative units; x , position of peak.

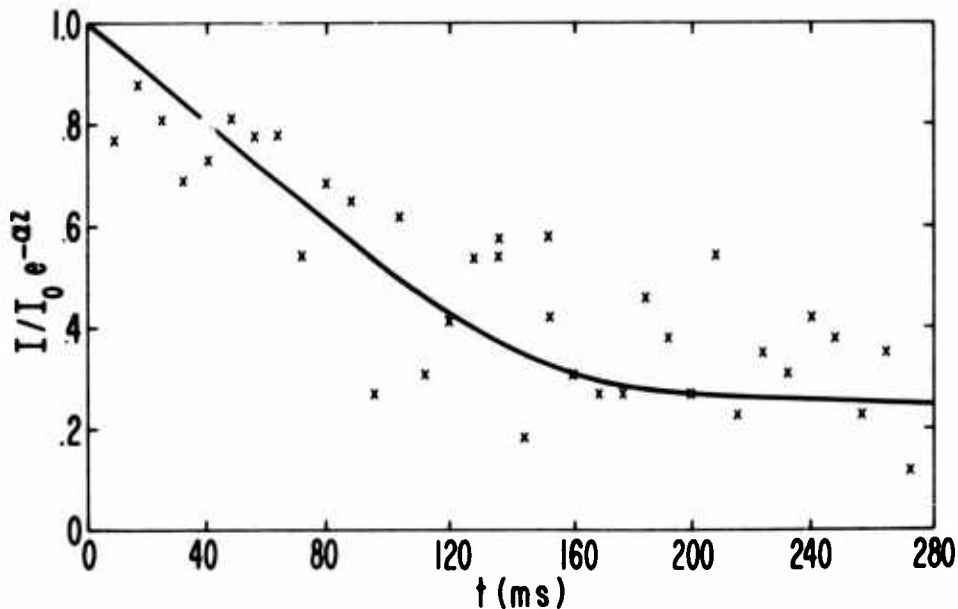


Fig. 13. Normalized beam intensity $N_f = f(t)$; $P = 100$ kW; $f = 80$ m; open atmosphere; — best fit to data.

A MORE RATIONAL APPROACH TO THE
STRESS ANALYSIS OF GUN TUBES

SHIH-CHI CHU, PhD
RESEARCH DIRECTORATE
GEN THOMAS J. RODMAN LABORATORY
ROCK ISLAND ARSENAL, ROCK ISLAND, ILLINOIS

On the basis of the Prandtl-Reuss stress-strain rule, the von Mises flow criterion, and the strain-hardening and compressibility properties of a material, an incremental theory is proposed in this paper for solving the problem of elastoplastic thick-walled tubes subjected to both mechanical and transient thermal loadings. In the developed method, incremental stresses and strains are directly used as variables; hence, numerical differentiation in the evaluation of stresses and strains is not required. Moreover, since no assumption is made on the displacement field, this method can provide better results than those of the finite-element methods. Since the consideration of stress, strain, loading, and temperature history is involved in the analysis, the present theory is particularly suitable for predicting stress and strain distribution, and location of elastic-plastic boundaries of a thick-walled cylinder such as a gun tube subjected to nonproportionate mechanical and transient thermal loadings.

INTRODUCTION

An important problem encountered in an effective design of gun barrels is the prediction and prevention of structural failure caused by internal pressure and transient thermal loading. To the investigator's knowledge, a rational methodology for the inelastic stress analysis for gun tubes is lacking at the present time. In the "Engineering Design Handbook for Gun Tubes," (AMCP 706-252), the use of Lamé's equations for open-end, thick-walled cylinders is recommended "guidance" in the determination of barrel wall thickness to withstand the stresses of firing. That approach, in which the entire barrel is considered to be loaded within the elastic region, results in

unrealistically large wall-thickness requirements even under conditions for which the room temperature mechanical properties of the gun barrel materials are applicable. For elevated temperatures, at which the strength of material is severely reduced, Lamé's approach would theoretically require a barrel with an extremely thick wall to resist the firing stresses. The current handbook does not provide the designer with an analytical solution to this important gun barrel design problem.

A literature survey indicated that the problems of elastoplastic thick-walled cylinders have been solved by almost all investigators on the basis of cylinders subjected to mechanical loadings [1-14]. Very little work has been done on the elastoplastic solutions for thick-walled cylinders subjected to transient thermal loadings which is very important for successful gun barrel design. The equivalent stress in a thick-walled tube under pressure and axial loading is a monotonic function with respect to the radius; hence, only one elastic-plastic boundary exists. However, for the problem of an elastoplastic thick-walled tube subjected to transient thermal loading, the equivalent stress in the cylinder is not necessarily a monotonic function with respect to the radius, but is strongly dependent upon the heat flow in the wall of the tube. Therefore, the number of elastic-plastic boundaries in the wall of the tube is not necessarily restricted to one, but may be two or more appearing at several places in the tube. Also, the temperature distribution, temperature gradient, and temperature history of the cylinder are very important factors in determining the locations of the elastic-plastic boundaries. Therefore, the problem of an elastoplastic thick-walled cylinder, such as a gun tube subjected to both mechanical and transient thermal loading, is much more difficult and more complicated than the problem of a thick-walled cylinder subjected only to mechanical loading. Moreover, no previous solution is available for this problem.

Bland [17] analyzed the problem of a thick-walled tube subjected to pressure and a steady-state temperature gradient. On the basis of the Tresca yield criterion, the incompressibility of a material, and the plane-strain assumption, Weiner and Huddleston [18] proposed a thermal stress analysis of heat-treated cylinders. By use of von Mises' yield criterion, Prandtl-Reuss' flow rule of plasticity, isotropic-isothermal hardening rule, and compressibility of a material, Chu [16] proposed a numerical thermoelastoplastic solution of a thick-walled tube. In all of the analysis above, the thermal stresses were induced by thermal expansion. However, the temperature effect on the yield strength of a material was not considered. Hence, these solutions can be applied to only the case in which the temperature in the material is lower than certain values. The yield strength of a

material decreases with temperature increases, and it decreases dramatically if the temperature of the material reaches a certain high level. At the present time, no solution is available for the elasto-plastic problems of thick-walled tubes subjected to both mechanical and transient thermal loading with the consideration of temperature yield strength of a material and nonisothermal flow rules of plasticity. The purpose of this investigation was to solve this complicated problem encountered in gun tube stress analysis.

DEVELOPMENT OF AN INCREMENTAL THEORY

In the development of an incremental inelastic theory for a thick-walled tube, a cylindrical coordinate system (r, θ, z) is used with the z axis coincident with the axis of the tube. The cross section of a thick-walled tube is divided into n rings by radii $r_1=a$, $r_2=\dots$, $r_k=c$, \dots , $r_{n+1}=b$, where $r_k=c$ is an elastic-plastic boundary.

For any point in the inelastic region, the governing equations for stresses, increments of stresses, and strains can be derived as follows:

Since Prandtl-Reuss' flow rule is assumed to be valid, two independent incremental stress-strain equations are derived as

$$\begin{aligned} & -\frac{1}{E}[(2\sigma_z - \sigma_r - \sigma_\theta) + \nu(2\sigma_r - \sigma_\theta - \sigma_z)]d\sigma_r + \frac{\nu}{E}[(2\sigma_z - \sigma_r - \sigma_\theta) - (2\sigma_r - \sigma_\theta - \sigma_z)]d\sigma_\theta \\ & + \frac{1}{E}[(2\sigma_r - \sigma_\theta - \sigma_z) + \nu(2\sigma_z - \sigma_r - \sigma_\theta)]d\sigma_z + (2\sigma_z - \sigma_r - \sigma_\theta)d\epsilon_r \\ & = (2\sigma_r - \sigma_\theta - \sigma_z)d\epsilon_z + \beta[(2\sigma_z - \sigma_r - \sigma_\theta) - (2\sigma_r - \sigma_\theta - \sigma_z)]dT \end{aligned} \quad (1)$$

where β is thermal expansion coefficient. All notations in the paper are the same as in Ref. (16), except those defined in the paper.

$$\begin{aligned} & \frac{\nu}{E}[(2\sigma_z - \sigma_r - \sigma_\theta) - (2\sigma_\theta - \sigma_z - \sigma_r)]d\sigma_r - \frac{1}{E}[(2\sigma_z - \sigma_r - \sigma_\theta) + \nu(2\sigma_\theta - \sigma_z - \sigma_r)]d\sigma_\theta \\ & + \frac{1}{E}[\nu(2\sigma_z - \sigma_r - \sigma_\theta) + (2\sigma_\theta - \sigma_z - \sigma_r)]d\sigma_z + (2\sigma_z - \sigma_r - \sigma_\theta)d\epsilon_\theta \\ & = (2\sigma_\theta - \sigma_z - \sigma_r)d\epsilon_z + \beta[(2\sigma_z - \sigma_r - \sigma_\theta) - (2\sigma_\theta - \sigma_z - \sigma_r)]dT \end{aligned} \quad (2)$$

The surface used to define the elastic limit is referred to as the yield surface. For strain-hardening nonisothermal material, the subsequent yield surface or loading function can be represented as

$$\begin{aligned}
& \left\{ \frac{1}{2\sigma} (2\sigma_r - \sigma_\theta - \sigma_z) + \frac{(1+\nu)}{E} \phi [(2\varepsilon_r - \varepsilon_\theta - \varepsilon_z) - \frac{(1+\nu)}{E} (2\sigma_r - \sigma_\theta - \sigma_z)] \right\} d\sigma_r \\
& + \left\{ \frac{1}{2\sigma} (2\sigma_\theta - \sigma_z - \sigma_r) + \frac{(1+\nu)}{E} \phi [(2\varepsilon_\theta - \varepsilon_z - \varepsilon_r) - \frac{(1+\nu)}{E} (2\sigma_\theta - \sigma_z - \sigma_r)] \right\} d\sigma_\theta \\
& + \left\{ \frac{1}{2\sigma} (2\sigma_z - \sigma_r - \sigma_\theta) + \frac{(1+\nu)}{E} \phi [(2\varepsilon_z - \varepsilon_r - \varepsilon_\theta) - \frac{(1+\nu)}{E} (2\sigma_z - \sigma_r - \sigma_\theta)] \right\} d\sigma_z \\
& - \phi [(2\varepsilon_r - \varepsilon_\theta - \varepsilon_z) - \frac{(1+\nu)}{E} (2\sigma_r - \sigma_\theta - \sigma_z)] d\varepsilon_r \\
& - \phi [(2\varepsilon_\theta - \varepsilon_z - \varepsilon_r) - \frac{(1+\nu)}{E} (2\sigma_\theta - \sigma_z - \sigma_r)] d\varepsilon_\theta \\
& = \phi [(2\varepsilon_z - \varepsilon_r - \varepsilon_\theta) - \frac{(1+\nu)}{E} (2\sigma_z - \sigma_r - \sigma_\theta)] d\varepsilon_z + \phi \beta dT + \frac{d\sigma(T)}{dT} dT
\end{aligned} \tag{3}$$

in which

$$\bar{\sigma} = \frac{1}{\sqrt{2}} [(\sigma_r - \sigma_\theta)^2 + (\sigma_\theta - \sigma_z)^2 + (\sigma_z - \sigma_r)^2]^{1/2} \tag{4}$$

$$\bar{\varepsilon}^p = \frac{\sqrt{2}}{3} [(\varepsilon_r^p - \varepsilon_\theta^p)^2 + (\varepsilon_\theta^p - \varepsilon_z^p)^2 + (\varepsilon_z^p - \varepsilon_r^p)^2]^{1/2} \tag{5}$$

and

$$\phi = \frac{2}{9} \frac{\eta E}{1 - \eta} \frac{1}{\bar{\varepsilon}^p} \tag{6}$$

where η is the strain-hardening factor for the material, $\sigma(T)$ is the temperature-dependent yield strength of the material, and ε_r^p is the plastic strain component in the r direction.

For any point in the elastic region, the Duhamel-Neumann stress-strain-temperature relations are assumed to be satisfied:

$$d\varepsilon_r = \frac{1}{E} d\sigma_r - \frac{\nu}{E} d\sigma_\theta - \frac{\nu}{E} d\sigma_z + \beta dT \tag{7}$$

$$d\varepsilon_\theta = -\frac{\nu}{E} d\sigma_r + \frac{1}{E} d\sigma_\theta - \frac{\nu}{E} d\sigma_z + \beta dT \tag{8}$$

$$d\varepsilon_z = -\frac{\nu}{E} d\sigma_r - \frac{\nu}{E} d\sigma_\theta + \frac{1}{E} d\sigma_z + \beta dT \tag{9}$$

The equation of compatibility and the equation of equilibrium are valid for both the elastic and the inelastic regions of a

thick-walled tube. The finite-difference forms of these two equations are given by [12]:

$$\begin{aligned}
 & - (r_{i+1} - r_i)(d\epsilon_r)_i + (r_{i+1} - r_i)(d\epsilon_\theta)_i + r_i(d\epsilon_\theta)_{i+1} \\
 & = (r_{i+1} - r_i)(\epsilon_r - \epsilon_\theta)_i - r_i[(\epsilon_\theta)_{i+1} - (\epsilon_\theta)_i]
 \end{aligned} \tag{10}$$

for the equation of compatibility, and

$$\begin{aligned}
 & (r_{i+1} - 2r_i)(d\sigma_r)_i - (r_{i+1} - r_i)(d\sigma_\theta)_i + r_i(d\sigma_r)_{i+1} \\
 & = (r_{i+1} - r_i)(\sigma_\theta - \sigma_r)_i - r_i[(\sigma_r)_{i+1} - (\sigma_r)_i]
 \end{aligned} \tag{11}$$

for the equation of equilibrium.

At each point $r=r_i$, six incremental quantities $d\sigma_r$, $d\sigma_\theta$, $d\sigma_z$, $d\epsilon_r$, $d\epsilon_\theta$, and $d\epsilon_z$ are present that have to be determined for each step of the variation of loading (this includes the variation of temperature distribution). Since the axial strain ϵ_z is independent of r , if the incremental of axial strain, $d\epsilon_z$, is specified in each increment of loads, then only five incremental unknowns are present at each point. Hence, a total of $5(n+1)$ unknowns exists that must be determined for each increment of loading. Five equations listed above can be formulated at each point (except $r=b$), either in the elastic region or in the plastic region. At the outer surface, $r=b$, of a thick-walled tube, some information concerning quantities in equations (10) and (11) is unavailable. Hence, the total number of equations is $5(n+1)-2$. To solve $5(n+1)$ unknowns, two additional equations resulting from the boundary conditions are:

$$(d\sigma_r)_{r=a} = -\Delta P_i \quad \text{and} \quad (d\sigma_r)_{r=b} = -\Delta P_o \tag{12}$$

SOLUTION PROCEDURE

To obtain a numerical elastoplastic solution of a thick-walled tube subjected to both mechanical and transient thermal loading, one must first determine the relationship of yield strength of the material vs. the temperature. Usually, the yield strength of a material decreases dramatically if the temperature reaches a certain high level. Test data of yield strength vs. temperature for chromium-molybdenum-vanadium (Cr-Mo-V) steel are shown in Figure 1 by open circles. A fourth-degree polynomial approximation for test data, as shown in Figure 1 by solid curve, is

CHU

$$\begin{aligned}\sigma(T) = & 131,000 - 10.89047T + 0.06483536T^2 \\ & - 0.0001666979T^3 + 0.00000005984565T^4\end{aligned}\quad (13)$$

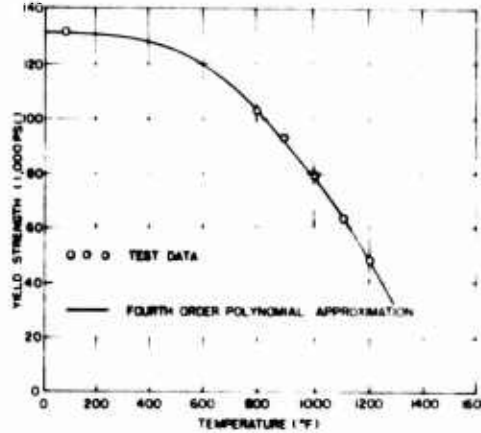


Fig. 1 Yield strength at elevated temperature for chromium molybdenum vanadium steel

This relation exhibits a very good approximation of test data.

To determine the transient thermal loading, one must first solve the transient temperature distribution. If the thermal properties of a material are assumed to be independent of the temperature, the heat flow in a thick-walled tube is governed by the well-known Fourier heat-conduction equation [19]

$$\frac{\partial^2 T}{\partial r^2} + \frac{1}{r} \frac{\partial T}{\partial r} = \frac{1}{\alpha} \frac{\partial T}{\partial t} \quad a < r < b, \quad t > 0 \quad (14)$$

where $\alpha > 0$ denotes thermal diffusivity. In addition, the following initial and radiation-convection boundary conditions are specified:

$$T(r, 0) = g(r) \quad a < r < b, \quad t = 0 \quad (15)$$

$$K \frac{\partial T}{\partial r} = -h_1(t)[T_g(t) - T_w] - \gamma F[T_g^4(t) - T_w^4] \quad r = a \quad (16)$$

$$K \frac{\partial T}{\partial r} = -h_2(t)[T_w - T_a(t)] - \gamma F[T_w^4 - T_a^4(t)] \quad r = b \quad (17)$$

Where $h_1(t)$ and $h_2(t)$ are the convection coefficients at the bore and outer surfaces, respectively, γ is the Stefan-Boltzman constant, K is thermal conductivity, and F is the interchange factor.

To provide all the calculations of the temperature distribution and the temperature history for this given problem, a finite-element digital computer code was developed [20]. For example, a three-layer compound barrel with insulating material in the middle layer under automatic firing condition is considered. The material properties of the barrel are given by:

	$\rho(\text{lb/ft}^3)$	$C(\text{Btu/lb-}^\circ\text{F})$	$K(\text{Btu/ft-hr-}^\circ\text{F})$
$.0495' \leq r < 0.0510'$	490	0.10	26.0
$0.0570' \leq r \leq 0.0597'$	36	0.25	0.1
$0.0597' < r \leq 0.0677'$	490	0.10	26.0

The duration of each shot cycle is considered to be 100 milliseconds. The variations of the propellant gas temperature, $T_g(t)$, and gas convection coefficient $h_1(t)$ are shown in Figure 2. $T_a=100^\circ\text{F}$, $T(r,0)=100^\circ\text{F}$, and $h_2=5 \text{ Btu/hr ft}^2\text{-}^\circ\text{F}$.

The bore temperature for the first five-shot cycles is shown in Figure 3. The maximum bore temperature is indicated at approximately 1.2 milliseconds after each firing.

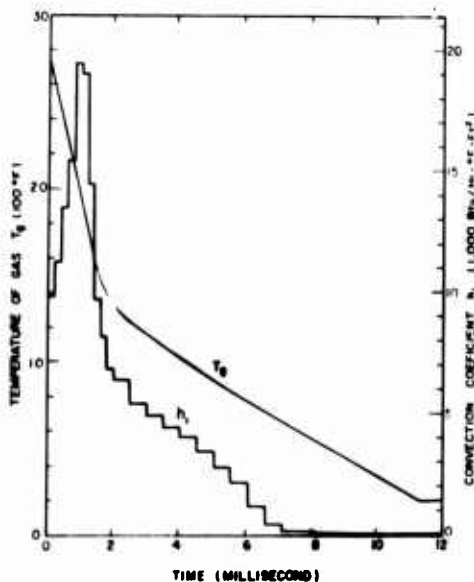


Fig. 2 Variations of propellant gas temperature and propellant convection coefficient

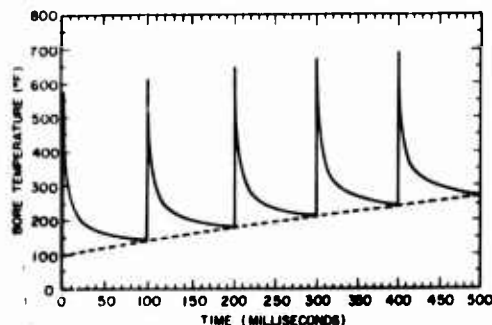


Fig. 3 Bore temperature history of a gun subjected to time dependent convection coefficient and radiation and time dependent propellant gas temperature

Since the stability and nonoscillation characteristics of the finite-element transient temperature solution were reported by Yalamanchili and Chu [21] in great detail, the discussion of instability and oscillation for the transient temperature numerical solution is omitted here.

The numerical computation procedure for obtaining thermo-elastoplastic solutions for thick-walled tubes can now be stated as follows: The computation starts with given loads (including thermal loading). The loading path is divided into a number of increments. At the beginning of each increment of loading, the distribution of stresses and strains is assumed to be known.

Step 1. Specify the values of $(d\sigma_r)_{r=a} = -\Delta P_i$, $(d\sigma_r)_{r=b} = -\Delta P_o$, and temperature variation $dT_i = \Delta T_i$ at each nodal point i .

Step 2. Assume a value for $d\epsilon_z$ (independent of r).

Step 3. Calculate $(d\sigma_r)_i$, $(d\sigma_\theta)_i$, $(d\sigma_z)_i$, $(d\epsilon_r)_i$ and $(d\epsilon_\theta)_i$ (for $i=1, 2, \dots, n+1$) from $5(n+1) \times 5(n+1)$ matrix equation formed by use of equations (1), (2), (3), (7), (8), (9), (10), (11), and (12).

Step 4. Compute $(\sigma_r)_i = (\sigma_r)_i \big|_{\text{before increase in loads}} + (d\sigma_r)_i$; $(\sigma_\theta)_i$, $(\sigma_z)_i$, $(\epsilon_r)_i$, $(\epsilon_\theta)_i$, and $(\epsilon_z)_i$ are computed in the same way.

Step 5. Compute the axial load, P_c , by use of Simpson's rule and compare it with the actual applied axial load P_a . If the difference between P_c and P_a falls within certain allowable limits, the computed stresses and strains in Step 4 are considered to be acceptable. Then, refer back to Step 1. Another set of new increments of mechanical load, variation of temperature distribution in the tube, and axial strain will be assigned to compute a new stress-and-strain distribution.

Step 6. If the difference between P_c and P_a is greater than some allowable limit, a new $d\epsilon_z$ has to be assumed. Then, go back to Step 2.

Step 7. The elastic-plastic boundary is determined by the condition that the effective stress $\bar{\sigma}$ is equal to the temperature-dependent yield strength, i.e.,

$$\bar{\sigma} = \frac{1}{\sqrt{2}} [(\sigma_r - \sigma_\theta)^2 + (\sigma_\theta - \sigma_z)^2 + (\sigma_z - \sigma_r)^2]^{1/2} \bigg|_T = \sigma(T) \quad (15)$$

In the elastic-plastic problem, the elastic-plastic boundary was allowed to be moved by an increment, at most equal to the thickness of a volume element.

For each step of the changing of mechanical loading or temperature distribution, a system of $5(n+1)$ linear algebraic equations, with nonzero terms clustered about the main diagonal, will be obtained. This type of matrix is known as a band matrix and can be solved quite rapidly on a digital computer. In the computer code that has been developed, the Gaussian elimination method is used to solve these equations.

APPLICATIONS

1. To illustrate the solution technique on the elastoplastic problem of a thick-walled tube subjected to both mechanical and thermal loading, a thick-walled tube is considered that has been subjected to internal pressure and heat input at bore surface with radiating and convecting boundary conditions on inner and outer surfaces. Without loss of generality, the initial temperature of the tube is assumed to be uniform with a value of zero. The radiation-convection boundary conditions were given in equations (18) and (19).

To obtain a numerical solution, the following values were assigned:

$\alpha = 4.3530 \text{ (ft}^2\text{/hr)}$, $\gamma = 0.1714 \times 10^{-8} \text{ (Btu/hr-ft-R)}$, $K = 219 \text{ (Btu/ft-hr-F)}$, $C_p \text{ specific heat} = 0.0915 \text{ (Btu/lb-F)}$, $F = 0.2$, $T_g = 3000^\circ\text{F}$, $T_a = 0^\circ\text{F}$, $h_1 = 100 \text{ (Btu/ft}^2\text{-hr-F)}$, $h_2 = 100 \text{ (Btu/ft}^2\text{-hr-F)}$, $b/a = 2.0$, $\nu = 0.3$, $E = 30 \times 10^6 \text{ psi}$, and $P_0 = 0$, and $\eta = 0.05$.

The cross section of the tube was divided into twenty volume-elements by radii $\rho_1 = 1.00$, $\rho_2 = 1.05$, \dots , $\rho_{21} = 2.0$. For each time increment Δt , the variation of temperature distribution for that particular time increment is computed, first by use of a finite-difference computer code; then this temperature variation and a increment of internal pressure ΔP_i are used as load inputs to compute all corresponding incremental stresses and strains by the method proposed in the previous section. For this particular example, plane strain assumption is considered to be valid. The distribution of stresses and strains, temperatures, and temperature-dependent yield strength at several times are given in Figures 4 through 7. The elastic-plastic boundary in the tube is located by use of equation (15).

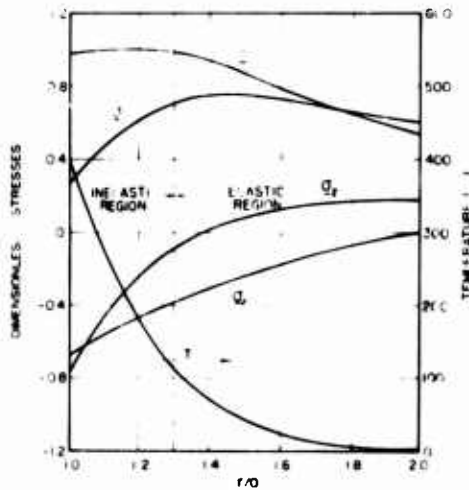


Fig. 4 Distribution of axial, radial, and circumferential stresses and temperature. ($t=61.30$ sec.)

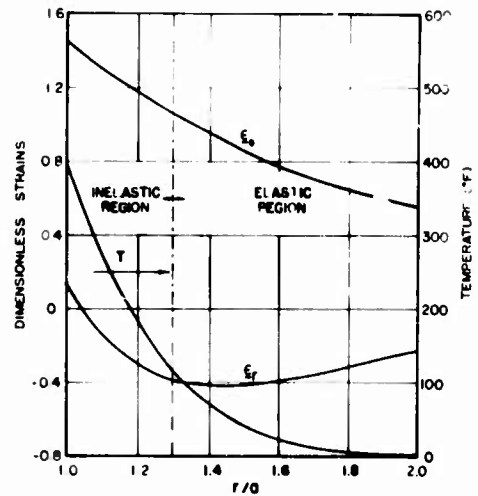


Fig. 5 Distribution of radial and circumferential strains and temperature. ($t=61.30$ sec.)

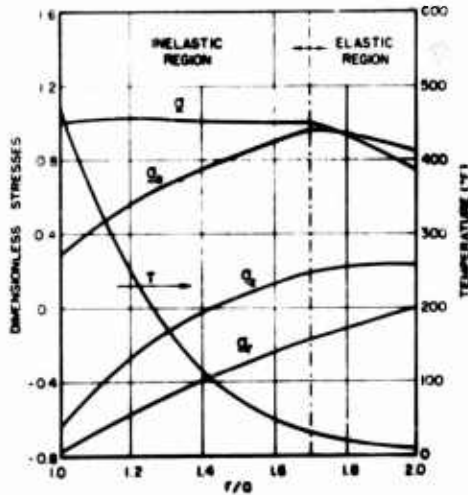


Fig. 6 Distribution of axial, radial, and circumferential stresses and temperature. ($t=90.80$ sec.)

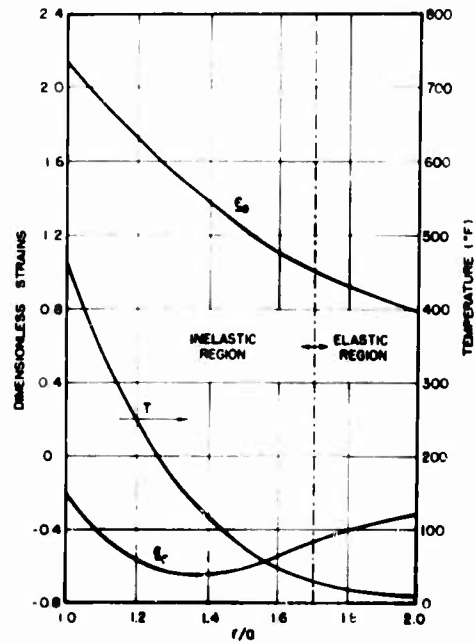


Fig. 7 Distribution of radial and circumferential strains and temperature. ($t=90.80$ sec.)

2. Another application of the present elastoplastic theory in determination of the wall ratio required at the breech end of a 7.62 mm barrel containing a peak proof-pressure of 70,000 psi is shown in Figures 8 through 10. These particular curves are based on the dynamic yield strength of Cr-Mo-V steel at a temperature of 1200°F, i.e., $\sigma_y \approx 109,000$ psi [22]. Both the Lamé' solution and the elastoplastic solution for maximum equivalent stress are shown for wall ratios of 3.0, 2.0, and 1.67 in Figures 8 through 10, respectively.

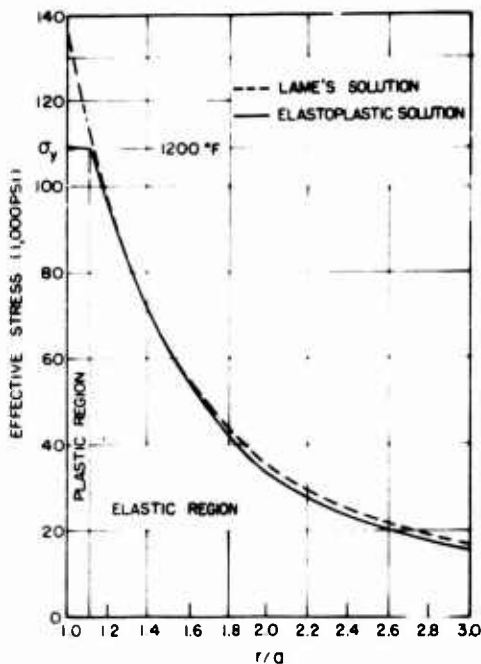


Fig. 8 Equivalent stress distribution for wall thickness ratio $W=3.0$ at an internal pressure of 70,000 psi (5% of the wall yielded)

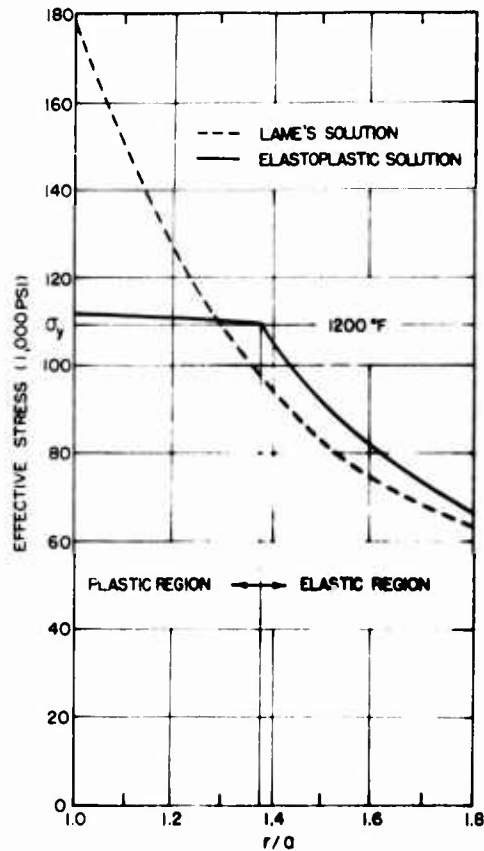


Fig. 9 Equivalent stress distribution for wall thickness ratio $W=1.8$ at an internal pressure of 70,000 psi (48% of the wall yielded)

From Figure 8, with Lamé's solution, a peak stress of 135,000 psi would be assessed at the inner fibers of a barrel having a breech end wall-ratio of 3.0. Since at 1200°F this stress exceeds the dynamic yield strength of 109,000 psi, failure would be predicted, i.e., classical gun tube design would require a much greater wall ratio. Conversely, with the elastoplastic solution, plasticity through only the inner 5 percent of the wall and a peak stress of slightly more than yield strength would be predicted. This is postulated as a conservative and completely safe condition. The increasingly excessive stresses predicted by Lamé's solution and the progression of plasticity into the barrel, under the elastoplastic theory (as the barrel wall ratio is reduced from 3.0 to 1.67), are shown in Figures 8 through 10.

The limiting wall ratio under the given conditions is shown in Figure 10 as having the approximate value 1.67 under the elastoplastic theory. This is the wall ratio at which the barrel should be almost completely plastic, but still capable of containing a pressure of 70,000 psi. With Lamé's solution, on the other hand, stresses of about 75 percent in excess of those that can be sustained by Cr-Mo-V steel at 1200°F would be predicted. Thus, with the elastoplastic theory, the resulting solutions are in much better agreement with the demonstrated capabilities of gun barrels.

3. The present theory is now being applied to a new 30 mm automatic weapon design. Since space is limited in this paper, the detail discussion of application of the present theory to this new weapon design is omitted here; however, it can be found in reference [23].

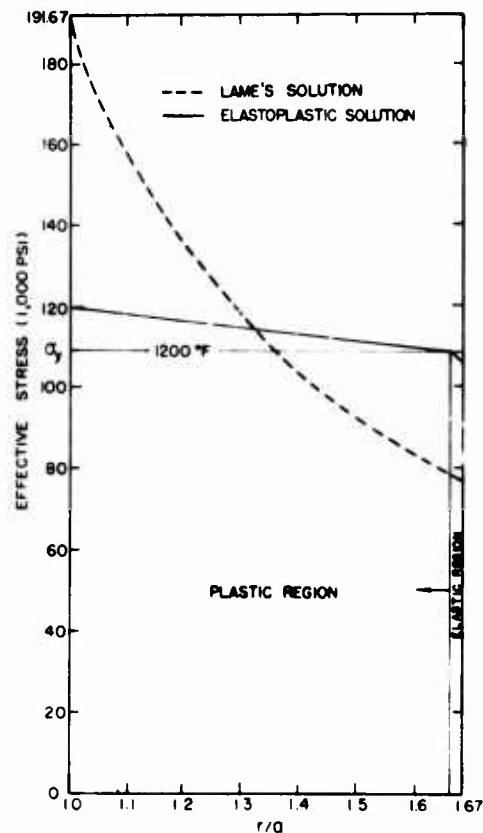


Fig. 10 Equivalent stress distribution for wall thickness ratio $W=1.67$ at an internal pressure of 70,000 psi

CONCLUSIONS

The increase in firing rate, firing capacity, and mobility of modern and future weapons has required that weapon structural components such as gun tubes, etc., be loaded near the limits of their load-carrying capacity. In the current "Engineering Design Handbook For Gun Tubes" (AMCP 706-252), the stress distribution is recommended that is calculated by Lamé's solution based on linear elasticity theory; hence, it results in unrealistically large wall-thicknesses of gun tubes. Thermal stresses are computed on the basis of the steady-state temperature distributions that are assumed to be logarithmic functions. However, before the steady-state temperature condition is reached, the transient temperature distribution in a gun tube is very important in the thermal stress analysis for gun tube design.

To obtain the most efficient design of gun tubes, a more rigorous nonlinear inelastic incremental stress analysis for gun tubes has been proposed in this paper. This is due to the necessity of designing for maximum stress to reduce weapon component weight and size to achieve an optimal weapon design. Instead of using the steady-state temperature distribution in a gun tube, the transient temperature distribution in a gun tube is computed in the present paper, on the basis of actual firing conditions of a gun. Since stress, strain, loading, and temperature history are considered in the analysis, the present theory is particularly suitable for predicting stress, strain, temperature distribution, and location of elastic-plastic boundaries of gun tubes subjected to nonproportionate mechanical and transient thermal loadings close to actual firing conditions in the field. The present theory is now applied to a new gun tube design.

On the theoretical point of view, the advantage of using the present theory over other theories can be noted as follows:

A literature survey indicated that the closed-form elasto-plastic solutions of thick-walled tubes [1,2,7,8] were obtained by use of deformation theory of plasticity. On the basis of the deformation theory, the strains at any state depend on the instantaneous state of stress and strain, and not on how that stress and strain system is reached. Taylor pointed out the inaccuracy of deformation theory, particularly for a member subjected to nonproportionate loading. More rigorous elasto-plastic solutions can be obtained only on the basis of the flow theory of plasticity. And, in such situations, one has to rely on numerical methods. Both the finite-element method and the finite-difference method are commonly used for solving elastoplastic problems.

To use the finite-element method [10,11] for solving elastoplastic problems, a functional has to be formulated on the basis of approximation expressions of the displacement field and the energy approach. Since the displacement field cannot be accurately determined at the beginning, the governing equations derived from the approximated functional do not exactly prescribe the inelastic behavior of a structural component. Hence, if a finite-element method must be used, some inherent errors cannot be avoided. The finite-difference treatment [3,5,6] of the elastoplastic problem of a thick-walled tube is usually based on the differential equations for the displacement vector and, hence, determination of stresses and strains requires numerical differentiation. However, the computer usually does not provide good results in differentiation unless a rather fine grid is used. With an alternative method developed in this investigation, incremental stresses and strains are directly used as variables and, hence, numerical differentiation in the evaluation of stresses and strains is not required. Moreover, since no assumption is made on the displacement field, this method can provide better results than those of the finite-element method.

REFERENCES

1. Cook, G., "The Stresses in Thick-Walled Cylinders of Mild Steel Overstrained by Internal Pressure," Proc. Inst. Mech. Engrs., Vol. 126 1934.
2. Sokolovsky, W. W., Theory of Plasticity, (in Russian and German), Chap. III, Gosudarstvennoe Izdatelstvo Tekhniko Teoretiicheskoi, Moscow, U.S.S.R., 1946.
3. Hill, R., Lee, E. H., and Tupper, S. J., "The Theory of Combined Plastic and Elastic Deformation With Particular Reference To A Thick Tube Under Internal Pressure," Proc. Roy. Soc. Lond., Ser. A., Vol. 1971, 1947.
4. MacGregor, C. V., Coffin, Jr., L. F., and Fisher, J. C., "Partially Plastic Thick-Walled Tubes," J. Franklin Inst., Vol. 245, 1948.
5. Hodge, P. C., and White, G. N., "A Quantitative Comparison of Flow and Deformation Theories of Plasticity," J. Appl. Mech. Vol. 72, 1950.
6. Hill, R., Lee, E. H., and Tupper, S. J., "Plastic Flow in a Closed-End Tube With Internal Pressure," Proc. First U. S. Nat. Cong. Appl. Mech., Chicago, Ill., 1951.
7. Allen, D. N., and Sopwith, D. G., "The Stresses and Strains in a Partly-Plastic Thick Tube Under Internal Pressure and End Load," Proc. Roy. Soc. Lond., Ser. A., Vol. 205, 1951.
8. Steele, M. C., "Partially Plastic Thick-Walled Cylinder Theory," J. Appl. Mech. Vol. 74, 1952.

9. Marcel, P. V., "A Note on the Elastic-Plastic Thick Cylinder With Internal Pressure in the Open and Closed-End Conditions," Int. J. Mech. Sci., Vol. 7, 1965.
10. Meijers, P., "Elastic-Plastic Deformation of Thick-Walled Cylinders," First Int. Conf. on Pressure Vessel Technology, Part I, Design and Analysis, ASME, 1969.
11. Chen, P. C. T., "The Finite Element Analysis of Elastic-Plastic Thick-Walled Tubes," Army Symposium on Solid Mechanics, Ocean City, Maryland, 1972.
12. Chu, S. C., "A More Rational Approach to the Problem of Elastoplastic Thick-Walled Cylinder," J. of Franklin Inst., Vol. 294, 1972.
13. Chu, S. C., and Chang, T. Y., "Inelastic Behavior of Thick-Walled Cylinders Made of Strain-Hardening Materials Subjected to Repeated Loading," Army Symposium on Solid Mechanics, Ocean City, Maryland, 1972.
14. Chu, S. C., and Vasilakis, J. O., "Inelastic Behavior of Thick-Walled Cylinders Subjected to Nonproportionate Loading," Experimental Mechanics, March, 1973.
15. Chu, S. C., "A Numerical Solution of an Elastoplastic Thick-Walled Tube Subjected to Thermal Loading," AIAA Paper, No. 73-256, Jan. 1973.
16. Chu, S. C., "A Numerical Thermoelastoplastic Solution of a Thick-Walled Tube," to be published in the AIAA Journal.
17. Bland, D. R., "Elastoplastic Thick-Walled Tubes of Work-Hardening Material Subject to Internal and External Pressures and to Temperature Gradient," J. of Mech. and Phys. of Solids, Vol. 4, 1956.
18. Weiner, J. H., and Huddleston, J. V., "Transient and Residual Stresses in Heat-Treated Cylinders," J. of Appl. Mech. Vol. 26.
19. Carslaw, H. S., and Jaeger, J. C., Conduction of Heat in Solids, 2nd Edition, 1959, Oxford.
20. Beckett, R. E., and Chu, S. C., "Finite-Element Method Applied to Heat Conduction in Solids With Nonlinear Boundary Conditions," Journal of Heat Transfer, Vol. 95, No. 1, Feb. 1973.
21. Yalamanchili, R. V. S., and Chu, S. C., "Stability and Oscillation Characteristic of Finite-Element, Finite-Difference, and Method of Weight Residuals for Transient Two-Dimensional Heat Conduction in Structures," Journal of Heat Transfer, May 1973, pp. 235-239.
22. "Gun Barrel Technology at Weapons Laboratory, Rock Island, 1968-1971," Brochure, Research Directorate, U.S. Army Weapons Command.
23. Thomsen, D. M., Chu, S. C., and Leech, W. J., "Gun Barrel Thermal Structural Model," Progress Report, Research Directorate, GEN Thomas J. Rodman Laboratory, Rock Island Arsenal.

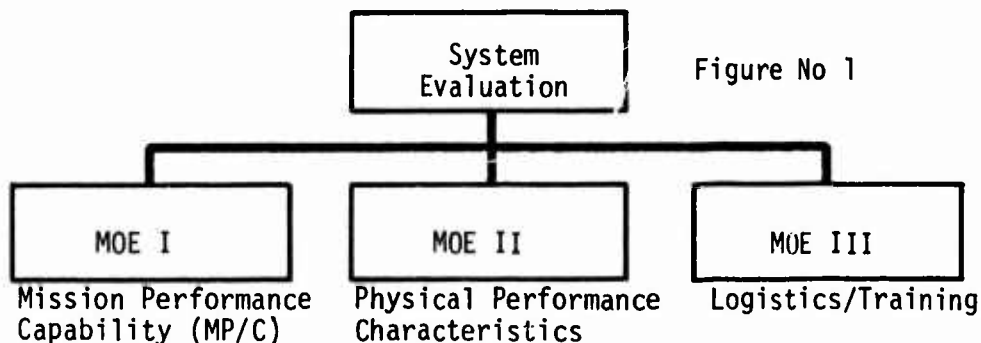
A METHODOLOGY/INSTRUMENTATION CONCEPT FOR TOTAL SYSTEM
TEST AND EVALUATION OF ARMY DEVELOPMENTAL MATERIEL

TERRELL G. COVINGTON, CPT, IN
ALLAN H. WEGNER, 1LT, AR
UNITED STATES ARMY ARMOR AND ENGINEER BOARD
FORT KNOX, KENTUCKY

INTRODUCTION

The development of this methodology/instrumentation concept for system test and evaluation evolved from and rests upon a dendritic structured evaluation/effectiveness schema. This schema was developed in conjunction with Armor and Engineer Board methodology studies on Measures of Effectiveness - a study to define the measures of effectiveness for armor and engineer test items which would undergo Development Testing Service Phase, Functional Field Testing - a study which developed the testing required for selected systems to adequately address data requirements for assessing the Mission Performance Capability, and Test Design - a study to define the general design of developmental tests.

The Measures of Effectiveness (MOE) study concluded that many armor and engineer test items exhibited the same general MOE of System Performance (Mission Performance Capability), Parametric Performance (Physical Performance Characteristics), and Logistics and Training.



COVINGTON & WEGNER

For specific systems, these MOE, and others when appropriate, are further subdivided to a level where field test measurement can be accomplished. (See incl 1.) The lowest level of the dendritic provides the data elements required for the evaluation of the materiel. For each data element, a scoring curve is derived based on criteria, items classified Standard A, and military experience. The entire dendritic can then be quantified to allow for aggregate analysis, as well as analysis of criteria.

The purpose of this manuscript is not to address all of the previously mentioned techniques but is instead intended to describe the design considerations for system testing (functional field testing) and its interface with instrumentation (XM56 Hit-Kill Indicator - SIMFIRE).

The recurring MOE, Mission Performance Capability, requires that the test item be immersed in its simulated employment environment and its mission performance capability be evaluated, the pure materiel system being tested as a totality to the maximum extent possible. This testing is referred to as functional field testing. In theory, the parametric criteria, if all met, will translate into a system performance which will adequately accomplish the intended mission. In fact, this may or may not be true, therefore, the tester must not only test against the criteria but must test the criteria as well. In addition to answering the mission performance capability issues per se, the functional field tests allow for the achievement of a perspective on the criterion/parametric performance of a system thus setting off system performance attained against the parametric performance and allowing the impact of ability or failure to meet criteria to be adequately addressed.

The following provides an example of the type evaluation indicated above: Through parametric/criterion testing, it is determined that a test tank's achievable top speed cross-country is 15 mph and its acceleration is 22 fps, while a standard or comparison tank achieves 12 mph and 17.6 fps respectively. The criteria requires 20 mph and 27 fps. Without system data, it is difficult, if not impossible, to address the impact of the test item's failure to meet the above criterion while surpassing the performance of a standard item. However, functional field testing addressing vulnerability/survivability might indicate that the maneuverability/agility accruing to the test tank owing to its increased achieved (over the standard tank) cross-country speed and acceleration, enables it to avoid being hit by enemy AT fire 50 percent more often than the standard tank. The significant increase in survivability of the test tank over the standard, despite failure to meet the parametric speed and acceleration criteria, augers for minimizing the impact of this failure and would argue strongly for considering this parametric insufficiency inconsequential. If further analysis were desired with respect to hit avoidance with parametric

performance at the criteria level, then this could be accomplished by computerized extrapolation or actual field testing by rigging a target vehicle with target and performance parameters equivalent to the tank and criteria respectively. It is recognized that cross-country speed and acceleration parameters also have fallout in areas other than vulnerability/survivability. However, they make their most direct, measurable, and quantifiable contribution to mission performance here; their other effects on such things as long-range movement rates of an armor unit being essentially muted out in a larger operational context.

FUNCTIONAL FIELD TEST DESIGN CONSIDERATIONS

There are certain general premises and principles which govern the application of functional field testing (FFT) techniques to any system. These methodology principles/guidelines provide the structural framework around which the particular functional test is designed. The actual design and substance of the FFT, as developed in accordance with the specified methodology concepts, then becomes a function of art and science. The ultimate test design must reflect the test developer's professional judgment and expertise as they relate to: definition of the mission spectrum and its reduction to essentials bearing upon the effectiveness of the tested system in its functional environment; development and precise definition of the data required to truly provide measures of the mission performance/capability of an item in the most comprehensive system context; identification and treatment of variables involved in the field testing; introduction of appropriate enemy action; and integration of mission reliability.

Whenever and wherever possible, a test materiel system should be evaluated in side-by-side testing with the standard item it is to replace, with competitive items or with proven systems of similar function/role. Comparison with a standard system is imperative if an effectiveness baseline or datum is to be established from which tested item effectiveness can be adjudged with perspective.

A. Mission Spectrum

Perhaps the most important consideration in test design is study, analysis, and development of the mission spectrum to be used in testing the item. The mission spectrum is absolutely pivotal to the entire test in that it dictates the demands to be placed on the system. If the mission spectrum and the associated demands are not sufficiently encompassing to include all critical demands, then the test may answer the wrong questions and omit vital considerations. The spectrum chosen must ensure inclusion of testing of the entire performance envelope of test and standard items. This is to assure that "additional capability" built into a new item is in fact tested

properly and the item allowed to benefit from it. For example, if a new missile firing is purportedly especially capable of engaging attacking tanks at long (3,000 + meters) ranges with great accuracy, then the scope of mission (FFT) should include this. Conversely, the scope must also include "attack" exercises to surface potential burdens introduced by addition of the long-range missile capability.

Very careful study and analysis must be given to ensuring that the demands associated with the mission spectrum focus upon the materiel system and the immediate operator/crew-system relationship to the exclusion of all else. Keeping this consideration in mind will also help reduce the actual mission spectrum to manageable proportions.

For instance in the case of the tank, though units to which it is assigned receive a gamut of missions ranging from penetration, envelopment, to delay actions, mobile defense ad infinitum - the individual tank system itself does not "care." For it, confrontations with other tanks obtain in three basic modes: attack against other defending tanks; defense from selected positions against other attacking tanks; and surprise confrontations with other tanks where the two come upon one another unexpectedly and neither side is either attacking or defending (meeting engagements). These three modes place the spectrum of critical demands on the materiel system, demands which do not change throughout the entire unit mission spectrum.

In the case of tactical bridging, care would have to be exercised to include in the mission spectrum a variety of gap configurations/spans, near and far bank orientations, weather and light conditions, traffic mix and duration and mission duration. This is necessary to demonstrate the expected MP/C of the bridge as a function of its flexibility in adapting to the interactive factors of gap span/configuration, load and traffic requirements. Specifically, a bridge's ability to permit tailoring to varying gap/load requirements and to accomplish the mission of passing traffic under varying mission requirements must be allowed to surface and be evaluated in conjunction with a comparison system.

Because of the peculiarities involving the MP/C assessment of support systems, what would normally involve FFT for evaluation of MP/C, can be accomplished via a model or computer simulation of the MP/C test environment. The model would use actual parametric performance test data to satisfy input-variable data requirements where necessary. Support systems are ideally suited to the model simulation approach for evaluation of MP/C because effects of enemy action can be effectively ignored or blocked out, and because the component variables of their MP/C are readily isolatable, few in number, and take the form of basic parametric characteristics which are readily tested and quantified in a test environment.

B. Mission Performance/Capability Measures

The finer measures of MP/C will vary with materiel types, though the single, broad measure of Degree of Mission Accomplishment (DMA) is applicable to MP/C of virtually all systems. As the name implies, DMA assesses the degree to which a materiel system accomplishes its intended purpose in execution of any given mission of the mission spectrum.

Because of the broad nature of the DMA measure, it is often desirable and judicious to provide a finer measure of, and an additional perspective on, MP/C.

In the case of the tank, DMA assesses the tank's ability to attack an objective or defend a position under various enemy-friendly tank force ratios. The basic element of the measure - mission accomplishment - addresses in go-no-go fashion, the tank's ability to either successfully attack and seize the objective destroying all enemy or to hold a position and destroy all attackers. The aspect of "degree" enters in the evaluation by addressing the extent to which the tank accomplished these attack/defend go-no-go missions over the entire mission spectrum; that is, out of a total of 16 attack missions under varying force ratios, the tank may accomplish 12. Obviously, because of the go-no-go nature of DMA, and its broad orientation, it may not alone provide the necessary discrimination between the performance of two different type tanks. A finer measure of the "efficiency" of MP/C is required in order to provide another perspective. Both type tanks may have accomplished 12 of 16 attack missions, yet one or the other may have suffered less losses and "killed" more enemy tanks in so doing. Hence, in the case of the tank, kill/loss ratios achieved in the missions provide a finer measurement and another perspective on MP/C by allowing the efficiency of MP/C to be addressed as well as the broader and basic issue of DMA.

A truck whose mission is to support the Class V (ammunition) requirements of a tank battalion participating in a variety of combat actions has its DMA assessed on the basis of whether or not it can supply the required Class V within the time constraints established. If the trucks do not deliver all the required Class V within the time parameters, they receive a zero for mission accomplishment. "Degree" is assessed in terms of the truck's ability to deliver Class V over the entire spectrum of combat action Class V requirements and the rapidity with which it could be delivered. Again, as in the case of the tank, DMA provides an insight into the truck's ability or inability to supply a required amount, but does not credit the truck for delivering a percentage of the requirement. Assessment of MP/C from the standpoint of cargo hauled as a percentage of that required, provides another evaluative slant which permits achievement of perspective on the actual cargo support capability, given a mission is not accomplished.

C. Identification and Treatment of Variables

Normally, the major and dominant variable in the FFT is the effectiveness or system performance of the tested materiel system, and it is this variable that is allowed to "run free" - then measured in the context of other variables/parameters which are either constan-tized, randomized, or blocked out completely.

The final test design for an item should represent the optimi-zation of the dynamic, but diametrically opposed interaction of maximi-zation of materiel parametric performance parameters allowed to have play in the mission spectrum of FFT and minimization of introduction of variables. The object is to maximize the system orientation of the test and mission realism while maintaining statistically significant and reliable results through minimization of variables and bias and the exercise of rigid control. The exercise of rigid control in the conduct of FFT of MP/C is vitally important to filter out leadership influences, tactical judgments, and other bias sources.

Attainment of a 100-percent assurance that a materiel system will not be penalized or profit by the bias sources must be a goal of FFT. A mechanism for achieving this is the employment of controllers with all test, standard, and control items in order to limit crew/operator responses and actions along predetermined lines.

In FFT, it is absolutely imperative that operator/crew sets be rotated between test and standard/comparison items with which the test materiel is being compared. This rotation tends to randomize and smooth the effect of variances in human ability. The greater the number of crew sets/operators that can be rotated through test and standard items during FFT, the greater the assurance that the results achieved will reflect the performance of the materiel when and if it is placed in the hands of representative users.

In order that the crew/operator skills be as representative of those to be found Army-wide, crew/operators for the test should be drawn from a variety of units/locations.

The design of FFT must ensure that tests generating data per-taining to assessment of MP/C are replicated sufficiently to allow statistical analysis of the data at statistically significant and re-liable levels. The specific number of replications required for a given test is a function of the number and nature of data element comparisons/assessments desired, the actual differentials in perform-ance demonstrated in testing, confidence levels of evaluations and the specific statistical analytical technique employed; hence, the exact number of replications can only be determined on a case-by-case basis.

The following is briefly illustrative of an approach to struc-turing variable interaction and control in designing a test for evalua-tion of the effectiveness of a minefield.

Effectiveness of a minefield has no meaning evaluated in a vacuum devoid of friendly-enemy interaction. This is true since any level of minefield potency can be achieved by increasing density if one is willing to expend the munition, personnel and time resources and cost to achieve it. A minefield is only effective/ineffective as a function of its contribution to friendly mission performance at an economically feasible level of employment/density. Consequently, the impact of the minefield on friendly mission performance becomes the operative variable which is allowed to run free, is measured, and the quantification of which becomes the measure of minefield effectiveness.

The operative variable of minefield effectiveness is then measured in a context of other randomized, constantized, and blocked out variables which are interwoven and constrained so that adequate statistical significance and reliability can be attached to results of testing. The infinite number of minefield variables of density and configuration (pattern, distribution, strips, depth) are first reduced to a spectrum or band of optimum/economically employable ones. The minefield density/configuration(s) employed in actual testing is then constantized at one or randomized using configurations from the reduced spectrum. A spectrum of terrain/vegetation conditions is determined and their influence randomized. An enemy-friendly force structure is selected based upon empirical data and constantized for the test. If test time/effort permit, these force structures may be randomly varied within a band of ratios; however, such an approach increases the number of replications required. Since probable enemy responses to a friendly force - minefield situation cannot be predicted reliably, the spectrum of relevant enemy responses is defined and reduced to those bearing upon minefield effectiveness. During testing iterations, these enemy responses are then varied randomly since they are basically unpredictable. Actual enemy and friendly movements, techniques, actions of crews, routes followed, etc., would be predetermined and constantized to preclude introduction of bias resulting from tactical judgment and leadership influences. The sizes of enemy and friendly forces involved in the evaluation is arrived at through a trade-off between the minimum force level at which breaching aids would be available and employed, accepted tactics are employable, and the maximum force levels which can be controlled to the rigid extent necessary and which the economics of sample size and cost make possible. The ideal compromise exists when an increase in enemy and friendly force levels would only have a linear effect on results due to the cellular nature of the enemy-friendly interaction, tactical techniques, etc.

Once the basic set of tests has been determined predicated on the minefield variables, the effort/time is compared to resources available to determine if they are realizable. If either time or materiel/personnel resources preclude the intended design, then the

number and band width of variables introduced (force ratios, terrain types, etc.) must be traded-off against effective test design with a view toward reducing the number of tests. Beyond a mere trade-off of variables, limited worst/best case testing may enable identification and elimination of non-impacting individual variables and reduction of other variable "bands" such as mine density, terrain types, etc.

D. Introduction of Enemy Action

Test and evaluation of combat systems under conditions of enemy activity is the keystone of FFT of those systems. The missions of combat systems exist only as they relate to performance of a function(s) against or in connection with an enemy force or activity. A priori, that is why a combat system is what it is.

Tests of combat systems in the absence of enemy action are ultimately sterile. They can only produce results obtained in a testing vacuum, results which say something about the tested item in and of itself - parametrically or system-wise, but nothing about how these characteristics or performance parameters are muted or operative in the face of enemy action in the real world. Two different type tanks could exhibit 63 and 76 percent hit probabilities on a range against standard targets; yet, in the face of enemy action (non-live fire, instrumented duplication of actual firing) in multi-iterative enemy-friendly mission context engagements, not evince significant differences in MP/C. This could develop because other system impacts pursuant to target acquisition and engagement under conditions of enemy action could attenuate the parametric difference in hit probabilities.

The only arena in which MP/C can be truly tested in light of enemy action is combat. However, the constraints of the test environment only allow approximation of the combat environment as a limit. The realism of enemy action can be affected in varying degrees either through realistic target presentations in a live-fire context or through the use of equipment which allows opposing force interaction in a non-live fire setting. This equipment consequently must require performance of crew/operator duties while duplicating the performance capabilities of the armament.

Introduction of such realistic enemy-friendly interaction into the testing environment at the USAARENBD is made possible through the use of hit-kill indicators based on the commercially available SIMFIRE system. The SIMFIRE system as manufactured by the British corporation Solartron and modified by its United States representative, EMR Telemetry, is presently being considered for adoption by the Army as a training aid, designated as XM56, Hit-Kill Indicator. SIMFIRE is an electro-optic device which duplicates the ballistics of actual tank gun munitions. Since the ballistic characteristics of actual ammunition is simulated, the crew must apply both the proper superelevation

and lead angles in addition to the proper sight picture to ensure a hit on the target. These three conditions are met only when all of the members of the crew perform their respective duties in the proper manner. Additionally, the target tanks are similarly equipped with SIMFIRE and hence able to "kill" the attacking tanks in return. Urgency motivation is thereby introduced into the exercise through the desire "to get him before he gets you." This sense of urgency, coupled with the fact that virtually all of the crewmembers must perform their crew duties in a precise fashion, stresses the man-machine system to its maximum in a quasi-combat environment.

SIMFIRE's main feature, its ability to ensure that all of the crewmembers must perform their respective, pertinent fire control duties quickly and accurately, is the basic feature needed in an instrumentation package to test combat vehicles in their intended mission roles. However, in the form it is commercially available, SIMFIRE only duplicates a few main armament systems, specifically, the British Chieftan Tank and the US M60A1 Tank. Because of this and other limitations, the basic SIMFIRE system had to be modified to duplicate a spectrum of direct-fire weapons mounted on a variety of vehicles, many of which are yet unknown. It is the modified form of SIMFIRE which is of interest in this paper, modifications which treble the original amount of electronics.

The additions to the basic SIMFIRE system are housed in one box and consist of the additional electronics needed to duplicate any tank main gun, antitank gun, any automatic weapon such as a 20MM automatic cannon, and any command guided missile such as TOW; in fact, any known direct fire weapon. Additional improvements to the original system include a rate sensing gyro to automatically calculate the proper lead angle and a probability of hit calculator to account for the numerous second order corrections that must be applied to the flight of a projectile. Because of safety considerations, high powered ruby lasers used in laser rangefinders (LRF) of many tanks cannot be used in two-sided testing, so provision was made to use the ranging capabilities of SIMFIRE as a substitute for those lasers. This enables the full capabilities of vehicles equipped with LRF's to be realized in exercises.

The modified SIMFIRE in its basic tank main gun/antitank gun configuration requires the following actions from the crew as appropriate. The tank commander first gives a fire command to coordinate actions of the crew. He then ranges to the target (using either LRF or coincidence rangefinder as appropriate to model). The driver will have to come smoothly to a stop. The gunner must index the proper ammunition in the computer and obtain the proper sight picture to include a proper lead if needed. Additionally, all of the fire control systems must, of course, be turned on and properly zeroed. The loader pushes a button to simulate loading of the round announced by tank

commander and switches the gun from the safe to fire position. While the loader does not handle an actual round, he must be an active member of the crew, and the system has a built-in time delay to simulate the loading of a round. If all of the above is done properly, the tank will score a hit on the target, dependent, of course, on the dispersion characteristics of the ammunition/weapon system which is accounted for in the manner previously mentioned. In any case, the gunner will receive fall of shot information from the SIMFIRE system to enable him to re-lay on the target for a second shot, if necessary. The above only happens if the opposing vehicle does not get him first and disable his system, or if he is not out of ammunition.

At the heart of the SIMFIRE system is GaAs laser emitting radiation at approximately 9000 Å with a peak power of approximately 10 watts. The beam from this laser then traverses a set of lenses which diverges the beam to a preset width. The width of the beam, as detected by the threshold detectors of the system, remains at approximately a constant width from about 200 M out to about 2500 M. This is accomplished by selecting the divergence of the beam to yield a power flow outward from the center of the beam to exactly compensate for the power flow outward from the detection threshold points to keep these points at a constant spacing as the beam propagates through space.

The beam is further made to scan a region of space around the target. If the target is in the central portion of the scanned zone, then the weapon is said to be "on target." However, if the target is not in the central portion of the scanned zone, the weapon is said to be off target.

The shape of the central portion of the scanned zone or "kill zone" has the shape of a quadrilateral with rounded corners, the size and shape of which may be varied to suit the vulnerable areas of the target. While the kill zone does not duplicate in general the exact area which a target might be vulnerable, this quadrilateral shaped kill zone usually is a good approximation. It is close enough in fact to require the attacking crew to perform their duties as if the target's vulnerable areas were exactly duplicated by the kill zone.

When engaging a target, SIMFIRE goes through three distinct phases. The first phase it goes into from the standby phase is the Ranging Phase. In this phase, the system ranges to the target. In the second phase or the Fall of Shot Phase, a determination is made where an actual round would impact. Where an actual round would impact with relation to the line of sight is determined from the range information of Phase 1 and the ballistics data of the round selected by the loader. If the weapon is "on target" the system then informs the target that it is "dead" during the third or Kill Phase. The first and third phases take 1/2 second each while the second phase takes 1 second for a total of 2 seconds for a complete engagement. In its modified version, provision is made to cut all of these times in half.

The gunner is then required to hold a good sight picture for only 1/2 second; not an unrealistic requirement.

To simulate a high-powered LRF, the modified SIMFIRE just cycles through Phase 1 and feeds the range information into the fire control circuits of the tank thus eliminating only the high-powered laser. All of the remainder of the fire control system remains intact and functioning which forces the crew to perform all duties to score a hit.

To provide a differential of performance between those vehicles which have a sophisticated fire control computer such as the M60A1E3 and a standard tank such as the M60A1, a probability of hit calculator is employed to account for those numerous second order corrections which are considered by these systems. However, when utilizing such a scheme, it is vital to recognize during a test when the assumption that the corrections are second order in nature is no longer valid. Such instances are, for example, when extremely high velocity crosswinds exist or unusually high cant angles exist. Fortunately, this shortcoming of the SIMFIRE system may be overcome through proper test design.

The next form of the modified SIMFIRE system to be considered is the command guided missile configuration. After being fired by the gunner, the system first ranges to the target, then goes into the Fall of Shot Phase and remains there until the calculated moment of impact. It then goes into the third phase and "kills" the target, if appropriate. During the second or Fall of Shot Phase, the ability of the gunner to keep the simulated missile in the allowed flight envelope to the target is continuously monitored. If he ever deviates from this region, he is credited with a miss. In the design of the missile configuration of the modified SIMFIRE, the elliptic paraboloid which describes the allowed flight cone to the target is described to the system by a piecewise linear approximation. Also, heavy reliance was placed on the observed fact that the average velocity and lateral acceleration of all of the known operational command guided missiles are highly correlated. This is probably due to the fact that as technology improved, both the average velocity and lateral acceleration showed a corresponding improvement.

Those instabilities in the flight of a missile induced by trying to track a dodging target, which are due to the intrinsic time delay of the human tracker to react to a change of apparent motion of a target, are similarly an inherent part of the human-SIMFIRE missile simulator. As such, no additional "electronics" is needed to duplicate this type of instability.

While the missile simulator is not the most highly sophisticated system, it, when coupled with the probability of hit calculator, provides a most acceptable first order simulation of the flight of an actual missile.

In its last configuration, the automatic weapon configuration, the system first ranges to the target after being fired by the gunner, then it goes into the fall of shot mode. It is in this mode where it is determined whether the gunner applied the proper superelevation and deflection to hit the target. The gunner is in receipt of continuous fall of shot information to aid him in his aim. Naturally, the gunner must have set the actual fire control switches in the vehicle including the rate control switch, if present, to their proper positions before he engages the target. The natural dispersion of automatic weapons is accounted for by the probability of hit calculator which takes the rate of fire as well as the range into account to determine if target has been successfully engaged. Because a single round from an automatic weapon is not likely to destroy a vehicle, the gunner must remain on target long enough for a sufficient, preset number of rounds to have hit the target. Additionally, since many automatic weapons have a maximum lethal range, the gunner must engage a target within this maximum, preset range.

This configuration is a logical extension of the basic tank gun SIMFIRE system except that more than one "fall of shot" determination is made to account for the multiple rounds fired by automatic weapons.

When considering any simulation system using an analog, it must be remembered that by their very definition they are imperfect duplications. These shortcomings of the simulation device must be recognized, and through the proper design of the overall test, any limitations of the simulation scheme may be minimized.

E. Integration of Mission Reliability

Mission reliability has its most realistic, proper, and most effective impact through introduction as a degrading function vis-a-vis MP/C. If merely evaluated separately in a vacuum and then set off against MP/C and other measures, its impact becomes difficult if not impossible to place in perspective.

In the case of combat systems, the immediate impact of mission reliability (mission failure) is similar to a loss resulting from enemy action. Consequently, an effective means of allowing mission reliability to exercise its full "play" impact on a combat system is to integrate it with the actual demonstration of MP/C as an attritional factor in loss rate complementing that resulting from enemy action.

In the case of support systems, mission reliability can be permitted to impact directly on MP/C by using it to "knock out" support systems during their mission execution, thereby, leaving less systems available to accomplish the total mission. The procedure for degrading MP/C with reliability performance is normally a paper process, after

testing is complete or sufficient reliability data gathered, and can be accomplished manually or by computer model depending upon the degree of sophistication of the FFT and evaluation.

RESULTS OF TECHNIQUES

The results from the use of the techniques discussed in this paper have been limited to early pilot tests involving the Add-On Stabilization System for M60A1 Tanks and the Mine Dispersing Subsystem for the XM56 Aircraft Delivered AntiTank Mine. Future major tests using these techniques include the Armored Reconnaissance Scout Vehicle and the M60A1E3 Tank Test.

In the M60A1 with Add-On Stabilization test, basic firing tests revealed that although the coaxial firing on the move was greatly improved, the main gun firing on the move was not as effective as had been hoped for. However, in SIMFIRE Functional Field Testing, although the test and comparison tank accomplished the mission in approximately the same number of instances, an important new piece of information was obtained. The test tank was able to minimize his time in a stationary position and was therefore "killed" only 16 times when fired on 52 times, while the comparison tank (M60A1) was "killed" 17 times in only 29 attempts. This information, not previously available, will give decision makers an improved perspective in determining if the improved capabilities outweigh the inherent increase in maintenance burden associated with an add-on device.

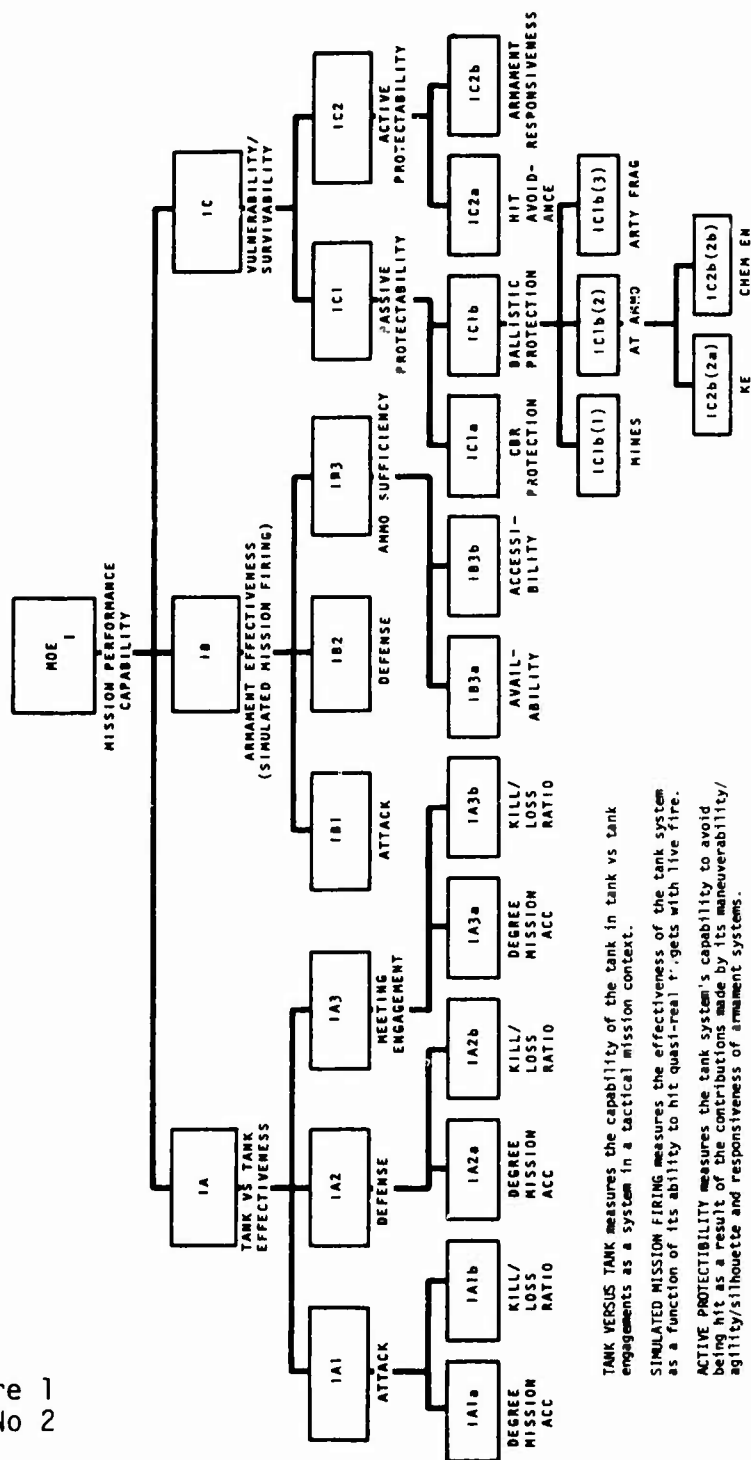
In the XM56 Mine Dispersing tests, functional tests involving tank vs tank engagements with and without a minefield, also provided previously unavailable information. In terms of the degree of mission accomplishment, it was found that a defending force was able to accomplish the defensive mission 18.8 percent of the time without the use of the minefield. However, with the use of a representative minefield (XM56) the defender was able to accomplish the defensive mission 62.5 percent of the time. It was also found that with the minefield in place, friendly losses were decreased by 29 percent and enemy losses were increased by 21 percent. These results were achieved with relatively low density, economical minefields. In previous mine testing, the individual mine was tested and evaluated. With functional field tests and SIMFIRE instrumentation, the minefield can be tested and evaluated in its employment environment.

From these results and the anticipated results of future tests, it is clear that this methodology/instrumentation concept will greatly enhance the evaluation of developmental materiel and will likewise provide valuable information to decision makers at all levels.

REFERENCES

1. USAARENBD Final Report of Methodology Investigation for Measures of Effectiveness, TECOM Project No 9-CO-00C-000-007, 29 Aug 73.
2. Test Operating Procedures for Functional Field Testing, TECOM TOP No 2-3-506, pending distribution.
3. USAARENBD Final Report of Development Test II of Mine Dispersing Subsystem, Aircraft, XM56 (U), MAJ Leo Holland, 29 Aug 73, TECOM Project No 7-WE-900-056-003. (Classified report (C))
4. USAARENBD Final Report of Initial Production Test of Add-on Stabilization System for M60A1 Tank (U), CPT Timothy O'Neill, 11 Jan 74, TECOM Project No 1-VC-08F-060-014. (Classified report (C))
5. USAARENBD Test Plan of Development Test II (Service Phase) of Tank, Combat, Full-Track, 105mm Gun, M60A1E3, MAJ Rudy Holbrook, 15 Oct 73, TECOM Project No 1-VC-080-060-046. (Classified report (C))
6. USAARENBD Test Plan of Competitive Developmental Test (Service Phase) of Armored Reconnaissance Scout Vehicle, XM800, MAJ Harry Mullis, 6 Aug 73, TECOM Project No 1-VC-040-800-009.

Inclosure 1
Figure No 2



TRACKING RELIABILITY GROWTH

LARRY H. CROW, PhD
U. S. ARMY MATERIEL SYSTEMS ANALYSIS AGENCY
ABERDEEN PROVING GROUND, MARYLAND

1. INTRODUCTION

Invariably, development programs for sophisticated, complex systems require considerable resources such as time, dollars and manpower, to achieve a level of system reliability acceptable to the user. The reliability requirements for many systems are high, and to obtain these high goals it is common practice to subject the system to a test-fix-test-fix process. During this process, the total system or major subsystems are tested to failure, system failure modes are determined, and design and/or engineering changes are made as attempts to eliminate these modes or, at least, to decrease their rate of occurrence. If this process is continued, and design and engineering modifications are made in a competent manner, then the system reliability will increase.

It is advantageous, of course, for the program manager to track this increase in system reliability during the development program. He may then determine as early as possible whether or not the system reliability is growing at a sufficient rate to meet the required goal and allocate available resources accordingly. In this regard, a program manager wishes to determine from test data the current reliability status of the system, estimate the rate of growth, and obtain projections of future expected reliability.

Since the system configuration is continually changing under this test-fix process, there is usually limited test data available on the system for a fixed configuration. Consequently, direct estimates of system reliability for a fixed configuration would generally not

Preceding page blank

CROW

enjoy a high degree of confidence and may, therefore, have little practical value.

Because of these difficulties with the direct estimation of system reliability, mathematical reliability growth models are often employed. Most reliability growth models considered in the literature assume that a mathematical formula (or curve), as a function of time, represents the reliability of the system during the development program. The central purpose of most reliability growth models includes one or both of the following objectives:

- a. Inference on the present system reliability,
- b. Projection on the system reliability at some future development time.

Many reliability growth models are parametric. That is, these models have certain parameters which are unknown and must be estimated from test data generated during the development program. This paper considers a popular parametric reliability growth model which is widely used in government and industry. Background on the derivation of the model will be discussed along with some major drawbacks with a "common sense" technique for estimating the unknown parameters. We show how these drawbacks can be avoided by applying estimation, goodness of fit and confidence interval procedures developed at AMSAA. Recently developed tables for computing exact confidence intervals on system failure rate and MTBF are given and an application of these techniques to an actual Army development program is discussed.

2. THE WEIBULL RELIABILITY GROWTH MODEL

In 1962, J. T. Duane of General Electric Company's Motor and Generator Department [see Duane (3)] published a report in which he presents his observations on failure data for five divergent types of systems during their development programs at G. E. These systems included complex hydromechanical devices, complex types of aircraft generators and an aircraft jet engine. The study of the failure data was conducted in an effort to determine if any systematic changes in reliability occurred during the development programs for these systems. His analysis revealed that for these systems, the observed cumulative failure rate versus cumulative operating hours fell close to a straight line when plotted on log-log paper. Similar plots have been noted in industry for other types of systems, and by the U. S. Army for various military weapon systems during development

CROW

[see Crow (2)].

For a mathematical interpretation of these straight line plots on log-log paper, let $N(t)$ denote the number of system failures by time t , $t > 0$. The observed cumulative failure rate $C(t)$ at time t is, therefore, equal to $C(t) = N(t)/t$. The plots on log-log paper imply that $\log C(t)$ is approximately a straight line. That is, $\log C(t) = \delta + \gamma \log t$. Equating $C(t)$ to its expected value and assuming an exact linear relationship, we have $\log (E[C(t)]) = \delta + \gamma \log t$. Taking exponentials gives $E[C(t)] = \lambda t^\gamma$, $\lambda = e^\delta$. Hence, $E[N(t)] = \lambda t^\beta$, for $\beta = \gamma + 1$, since $E[C(t)] = E[N(t)]/t$. Thus, the expected number of system failures by time t is λt^β .

The instantaneous failure rate, $r(t)$, of the system is the change per unit time of $E[N(t)]$. Thus, $r(t) = \frac{d}{dt} E[N(t)] = \lambda \beta t^{\beta-1}$, which is recognized as being the Weibull failure rate function. It is important to note that since the system configuration is changing, the data are not homogeneous and, therefore, the usual theory for a Weibull distribution will not apply. In fact, it has been shown by the author [see Crow (1)] that when the configuration of the system is changing, and failures are governed by the failure rate $r(t) = \lambda \beta t^{\beta-1}$, then the system failure times follow a nonhomogeneous Poisson process with Weibull intensity function $r(t)$.

At time t_0 the Weibull failure rate is $r(t_0) = \lambda \beta t_0^{\beta-1}$. If no further system improvements are made after time t_0 , then it is reasonable to assume that the failure rate would remain constant at the value $r(t_0)$ if testing were continued. In particular, if the system were put into production with the configuration fixed as it was at time t_0 , then the life distribution of the systems produced would be exponential with mean time between failure (MTBF) $M(t_0) = [r(t_0)]^{-1} = t_0^{1-\beta}/\lambda \beta$. Hence, for $0 < \beta < 1$, the MTBF $M(t)$ increases as the development testing time t increases, and is proportional to $t^{1-\beta}$. Thus, β is a growth parameter reflecting the rate at which reliability, or MTBF, increases with development testing time.

If this Weibull model is determined to sufficiently represent the occurrence of failures for a particular system during development testing, then it can, of course, be used to monitor and

CROW

project the growth of system reliability. To do this, however, would require estimating from test data the two unknown parameters λ and β by say $\tilde{\lambda}$, $\tilde{\beta}$. One would then estimate the failure rate function by $\tilde{r}(t) = \tilde{\lambda}\tilde{\beta}t^{\tilde{\beta}-1}$ and the MTBF function by $\tilde{M}(t) = [\tilde{r}(t)]^{-1} = t^{1-\tilde{\beta}}/\tilde{\lambda}\tilde{\beta}$. If the system is tested to time T , say, then $\tilde{M}(T)$ would estimate the current MTBF, and $\tilde{M}(t)$, $t > T$ would project estimates of system MTBF into the future.

Consider a "common sense," often used procedure for estimating λ and β . Suppose the system is tested to time T , and let $0 < T_1 < T_2 < \dots < T_K = T$ be a partition of $(0, T]$. The observed cumulative failure rate at time T_i is $C(T_i) = N(T_i)/T_i$, where $N(T_i)$ is the number of system failures to time T_i , $i = 1, \dots, K$. Recall that $\log E[C(T_i)] = \log \lambda + (\beta-1)\log T_i$. Hence, if we plot $\log C(T_i)$ versus $\log T_i$ on coordinate paper and fit a line by linear regression, we could use γ , the slope, to estimate $\beta-1$ and δ the intercept at $t = 1$ to estimate $\log \lambda$. The estimates of λ and β would be $\tilde{\lambda} = e^{\delta}$, $\tilde{\beta} = \gamma + 1$, respectively.

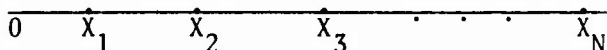
There are several points to be made about the above techniques for estimating λ and β . Firstly, the estimates are dependent on the choice of T_i , $i = 1, \dots, K$, and, of course, may differ for different choices. Thus, this method is subjective, yielding results perhaps not susceptible to rigorous analysis. Secondly, the values $C(T_i)$, $i = 1, \dots, K$ are not independent since $N(T_i) \leq N(T_j)$ for $i < j$. Moreover, the variances of the $C(T_i)$'s are not equal. In particular, $\text{Var}[C(T_i)] = \lambda T_i^{\beta-2}$. If the system reliability is improving ($0 < \beta < 1$), then $\text{Var}[C(T_i)]$ is decreasing as T_i increases. Hence, since the $C(T_i)$'s are not independent with equal variances, usual normal regression theory will not apply to yield confidence bounds on the parameters λ , β , and the functions $r(t)$, $M(t)$. Finally, in practice, the criteria for using the Weibull model and this estimation technique would probably depend on the subjective appraisal of whether or not the plotted points appear to lie nearly on a straight line.

It is apparent that improved goodness of fit, estimation and confidence bound procedures are needed for this highly important task of monitoring and projecting the growth of system reliability during development. Using the result that the plots on log-log paper imply that the successive failure times of the system follow a certain stochastic process (i.e., the nonhomogeneous Poisson process with Weibull intensity $\lambda \beta t^{\beta-1}$) we have derived a variety of useful statistical procedures for this model. Some recent results will be discussed in the following sections.

3. ESTIMATION AND GOODNESS OF FIT PROCEDURES

If the successive times of failures are being recorded for a system undergoing development testing, then a statistical goodness of fit test can be performed to determine if the Weibull reliability growth model is appropriate. If the model is acceptable, then closed form maximum likelihood (ML) estimates of λ and β may be used to estimate and project system MTBF. Using these procedures developed by the author in (1), one can avoid the aforementioned drawbacks associated with estimation from log-log plots.

Suppose that a system has experienced N failures during development testing. Let X_i be the age (time on test) of the system at the i -th failure $i = 1, \dots, N$. If testing is stopped at the N -th failure time, the data are said to be failure truncated.



The ML estimate of β , the growth parameter, is

$$(3.1) \quad \hat{\beta} = \frac{N}{\sum_{i=1}^{N-1} \log \frac{X_N}{X_i}},$$

and the ML estimate of λ is

$$(3.2) \quad \hat{\lambda} = \frac{N}{\hat{X}_N^{\hat{\beta}}}.$$

Thus, calculating $\hat{\lambda}$, $\hat{\beta}$ one may estimate the failure rate function $r(t) = \lambda \beta t^{\beta-1}$ by $\hat{r}(t) = \hat{\lambda} \hat{\beta} t^{\hat{\beta}-1}$. The MTBF function $M(t) =$

CROW

$[r(t)]^{-1}$ is similarly estimated by $\hat{M}(t) = [\hat{r}(t)]^{-1}$. In particular, the current estimate of the MTBF is $\hat{M}(X_N) = X_N/N\hat{\beta}$, and $\hat{M}(t)$, $t > X_N$, projects expected future growth of system MTBF.

To determine the appropriateness of the Weibull model for representing the reliability growth for this system, one may calculate the statistic

$$(3.3) \quad C_M^2 = \frac{1}{12M} + \sum_{i=1}^M \left[\left(\frac{X_i}{X_N} \right)^{\bar{\beta}} - \frac{2i-1}{2M} \right]^2,$$

where $M = N - 1$, $\bar{\beta} = [(M-1)/N]\hat{\beta}$. Critical values of the C_M^2 statistic for $M = 2$ thru 60 have been determined at AMSAA from Monte Carlo simulation, using 15,000 samples for each value of M . Various critical values are given in Table 2 of (1).

If the statistic C_M^2 is greater than the selected critical value, then the Weibull model is rejected at the designated significance level. If C_M^2 is less than this critical value, then the Weibull model is accepted and may be used to track the system reliability growth.

Suppose that $K \geq 1$ systems have been simultaneously tested to time T , where T is not a failure time. In this case the data are time truncated. If design and engineering modifications are made on all K systems at the same time, then at any time during the testing the systems will have basically the same configuration. In this situation, we may combine the failure data on these K systems to obtain estimates of λ and β . These estimates and other related procedures are given in (2).

4. CONFIDENCE BOUNDS FOR MTBF

In this section we shall give recently developed procedures for placing confidence bounds on current and projected failure rates and MTBF. These procedures apply to the single system, failure truncated situation. Similar developments for time truncated testing will appear in a future AMSAA report when completed.

If a system undergoes development testing until the N -th

CROW

failure occurs, then $r(X_N)$ $[M(X_N)]$ is the current failure rate [MTBF]. It can be shown that the ratio $U_N = Nr(X_N)/(N-1)\hat{r}(X_N)$ is distributed independently of λ and β , where $\hat{r}(X_N)$ is the ML estimate $\hat{\lambda}\hat{\beta}X_N^{\hat{\beta}-1}$ of $r(X_N)$. Percentage points of this ratio were obtained at AMSAA from Monte Carlo simulation for $N = 2$ thru 60. These percentage points are presented in Table 1. Exact $100(1-\alpha)$ percent confidence bounds on $r(X_N)$ are of the form $[\hat{r}(X_N)a(N-1)/N, \hat{r}(X_N)b(N-1)/N]$, where a and b are from Table 1 such that $\text{Prob}(a < U_N < b) = 1-\alpha$. Equivalently, $100(1-\alpha)$ percent confidence bounds on $M(X_N) = [r(X_N)]^{-1}$ are of the form $([\hat{r}(X_N)b(N-1)/N]^{-1}, [\hat{r}(X_N)a(N-1)/N]^{-1})$.

For $N > 60$, $100(1-\alpha)$ percent confidence bounds may be calculated from the approximate relationships: $a \doteq 1 - \sqrt{2/N} Z_{\alpha/2}$, $b \doteq 1 + \sqrt{2/N} Z_{\alpha/2}$, where $Z_{\alpha/2}$ is the $\alpha/2$ -th percentile for the standard normal distribution.

For N moderately large, we may also use the percentage points in Table 1 to place approximate confidence bounds on future failure rates and MTBF. In particular, suppose we wish to place approximate $100(1-\alpha)$ percent confidence bounds on $r(T)$, $T > X_N$.

These approximate confidence bounds will again be of the form

$[\hat{r}(T)a(N-1)/N, \hat{r}(T)b(N-1)/N]$, where $\hat{r}(T) = \hat{\lambda}\hat{\beta}T^{\hat{\beta}-1}$ is the ML estimate of $r(T)$, and a and b are the appropriate percentage points from Table 1. Approximate $100(1-\alpha)$ percent confidence bounds on $M(T)$, the MTBF at time T , are derived, as before, from the bounds on $r(T)$. These bounds become exact as $N \rightarrow \infty$.

5. NUMERICAL EXAMPLE

Suppose that a system undergoing development testing recorded the following 40 successive failure times; .7, 3.7, 13.2, 17.6, 54.5, 99.2, 112.2, 120.9, 151.0, 163.0, 174.5, 191.6, 282.8, 355.2, 486.3, 490.5, 513.3, 558.4, 678.1, 688.0, 785.9, 887.0, 1010.7, 1029.1, 1034.4, 1136.1, 1178.9, 1259.7, 1297.9, 1419.7, 1571.7, 1629.8, 1702.3, 1928.9, 2072.3, 2525.2, 2928.5, 3016.4, 3181.0, 3256.3. That is, the system was of age .7 when the first failure occurred, of age 3.7 when the second failure occurred, etc. At age

CROW

3256.3 the system had the 40-th failure. From these data, and equations (3.1) and (3.2) we find that $\hat{\lambda} = 0.761$, $\hat{\beta} = 0.490$.

To determine if the Weibull model may be used to track this system's reliability growth, we calculate the goodness of fit statistic C_M^2 given by equation (3.3) where $M = 39$, $\bar{\beta} = (38/40)\hat{\beta} = 0.465$. This gives $C_{39}^2 = 0.077$. Next, we find in Table 2 of (1) that for $M = 39$, the critical value at the .05 significance level is 0.218. Since $C_{39}^2 < 0.218$, we accept the Weibull model.

Using $\hat{\lambda}$, $\hat{\beta}$, the failure rate function is estimated by $\hat{r}(t) = \hat{\lambda}\hat{\beta}t^{\hat{\beta}-1}$ and the MTBF function is estimated by $\hat{M}(t) = [\hat{r}(t)]^{-1}$. The current failure rate $r(3256.3)$ is estimated to be $\hat{r}(3256.3) = 0.006$, and the estimate of current MTBF is $[\hat{r}(3256.3)]^{-1} = 166.7$.

To place 90 percent confidence bounds on the current MTBF $M(3256.3)$, we refer to Table 1, $N = 40$, and find $a = 0.664$, $b = 1.40$. Using the formulas in the previous section, we get 90 percent confidence bounds (0.004, 0.008) for $r(3256.3)$. Hence, 90 percent confidence bounds on $M(3256.3)$ are (125.0, 250.0).

Suppose we wish to place approximate 90 percent confidence bounds on future MTBF, say at $T = 4000$. Using $\hat{r}(4000) = 0.005$, we calculate these bounds to be (0.003, .007). Approximate confidence bounds on $M(4000)$ are, therefore, (142.8, 333.3).

6. APPLICATION

In this section we shall discuss an application of the Weibull reliability growth procedures to an Army development program. Two major points concerning the application of this model are demonstrated. Firstly, the model may be applied to discrete data. Secondly, as in any mathematical model, care should be exercised in its use. In particular, the importance and usefulness of the goodness of fit statistic in Section 3 is demonstrated in this application.

Recently, AMSAA conducted a reliability growth study of a missile system. The purpose of the study was to use historic data on the first 801 valid flight tests to determine the growth curve, and

CROW

also to ascertain in retrospect how these data could have been used to track and project system reliability during development.

In reliability growth considerations, it is configuration changes on the system which are of prime importance. Consequently, in this study the 801 valid flights were ordered according to manufacturing date, since this should reflect the sequence and consequences of system configuration changes during development. The data consisted of the flight numbers at which a missile failure occurred. Observed that these are discrete data as opposed to continuous data in the model. However, it can be shown that for a large number of data points, the discrete failure process can be approximated by the continuous model. This approximation improves as the number of data points increases.

The interpretation of $r(t)$ for this type of application is that $r(i) = \lambda \beta i^{\beta-1}$ is the probability of failure for the i -th missile produced, $i = 1, 2, \dots$. Hence, $R(i) = 1 - r(i)$ is the reliability of the i -th missile. Analogous to MTBF, $M(i) = [r(i)]^{-1}$ is the mean flight between failure.

The first step in determining the reliability growth curve was to use the failure results for the 801 flights, and equations (3.1) and (3.2) to estimate the parameters of the Weibull model. The goodness of fit statistic C_M^2 , given by equation (3.3), was then calculated to determine if the model and data were compatible. The value of the statistic was highly significant (very large) indicating that the model did not reasonably represent the data. This implies that a single, smooth, Weibull curve would not reflect the decrease in failure probability of this system.

Further investigation revealed that the development program experienced a major re-emphasis on reliability improvement after the 200-th flight. Thus, the parameters of the model were estimated separately for the first 200 flights (see Figure 1) and for the remaining 601 flights (see Figure 2). In both cases, the goodness of fit of the model to the data was acceptable. The horizontal lines in Figures 1 and 2 are the average failure probabilities over 100 flight intervals. The smooth curves are the estimated Weibull failure

probabilities $\hat{\lambda} \hat{\beta} i^{\hat{\beta}-1}$. These curves are solid up to the end of the data, and the dash lines indicate the estimated future decrease in failure probability if the current rate of improvement were continued.

CROW

From the two curves the reliability $R(i) = 1 - r(i)$ is estimated. The resulting reliability growth curve is shown in Figure 3 with a jump at 200. The magnitude of the jump was calculated by parametric and nonparametric means, and consultation with the program office.

We next considered how the Weibull model could have been used to track and project system reliability during development. Using the first 200 flights, the estimate of the current reliability was .68 and the projected reliability at flight 800 was .74 (Figure 1). This projection indicated that the system reliability requirement would not be met if the present trend were continued. There was a major re-emphasis on reliability, and based on the next 100 flights (201-300), an estimate of the reliability at 300 was .89 and a projection to 800 was .94 (Figure 4). This projection was very close to the current estimate of .95 for system reliability obtained using all the data on flights 201-800 (Figure 3).

Thus, the estimation procedures provided a good guide as to when additional emphasis should be placed on reliability, and also provided accurate estimates of future system reliability for each phase of the development program.

ACKNOWLEDGEMENT

The author wishes to thank Mr. Edward F. Belbot for the computer programming which generated the Monte Carlo results of Table 1.

REFERENCES

1. L. H. Crow, Reliability Analysis For Complex, Repairable Systems, AMSAA Reliability, Availability and Maintainability Division Interim Note No. 26, January 1974.
2. L. H. Crow, On Reliability Growth Modeling, AMSAA Reliability, Availability and Maintainability Division Interim Note No. 27, January 1974.
3. J. T. Duane, Learning Curve Approach to Reliability Monitoring, IEEE Trans. Aerospace, Vol. 2, 1964.

N \ P	2.5	5.0	10.	20.	30.	40.	50.	60.	70.	80.	90.	95.	97.5
2	.013	.027	.059	.135	.230	.353	.511	.725	1.03	1.49	2.44	3.59	4.95
3	.065	.099	.165	.282	.401	.540	.698	.888	1.14	1.51	2.21	2.94	3.70
4	.125	.168	.243	.375	.496	.631	.784	.958	1.17	1.48	2.02	2.57	3.13
5	.174	.227	.308	.438	.559	.682	.834	.984	1.17	1.44	1.90	2.37	2.85
6	.218	.274	.357	.487	.606	.723	.845	.991	1.17	1.41	1.82	2.21	2.61
7	.249	.312	.393	.521	.636	.749	.869	1.00	1.17	1.39	1.76	2.09	2.44
8	.285	.343	.427	.554	.663	.773	.887	1.01	1.17	1.37	1.70	2.01	2.31
9	.310	.373	.459	.582	.690	.792	.899	1.02	1.16	1.35	1.64	1.91	2.20
10	.336	.401	.486	.607	.709	.805	.908	1.02	1.16	1.34	1.62	1.89	2.15
11	.361	.422	.508	.627	.726	.821	.920	1.03	1.15	1.32	1.58	1.84	2.08
12	.379	.441	.527	.644	.740	.827	.923	1.03	1.15	1.31	1.55	1.79	2.01
13	.397	.456	.542	.659	.751	.842	.931	1.03	1.15	1.29	1.53	1.76	1.98
14	.420	.478	.559	.667	.759	.847	.936	1.03	1.15	1.29	1.52	1.73	1.94
15	.427	.489	.574	.681	.772	.858	.942	1.03	1.14	1.28	1.50	1.71	1.90
16	.442	.505	.586	.690	.780	.861	.946	1.03	1.14	1.27	1.49	1.68	1.87
17	.456	.519	.596	.699	.787	.867	.949	1.04	1.14	1.27	1.47	1.65	1.83
18	.466	.523	.600	.710	.794	.873	.953	1.04	1.14	1.26	1.45	1.64	1.80
19	.479	.538	.613	.715	.801	.881	.957	1.04	1.13	1.26	1.44	1.62	1.78
20	.490	.548	.623	.723	.804	.881	.957	1.04	1.13	1.25	1.43	1.61	1.77
21	.498	.561	.634	.733	.810	.883	.958	1.04	1.13	1.24	1.42	1.59	1.74
22	.509	.566	.641	.736	.813	.887	.959	1.04	1.12	1.24	1.41	1.57	1.71
23	.511	.574	.645	.742	.820	.892	.963	1.04	1.12	1.23	1.40	1.55	1.69
24	.522	.578	.651	.746	.823	.892	.963	1.04	1.12	1.23	1.40	1.54	1.68
25	.528	.586	.659	.754	.826	.897	.966	1.04	1.12	1.22	1.38	1.52	1.65
26	.530	.588	.660	.756	.831	.900	.967	1.04	1.12	1.22	1.38	1.52	1.65
27	.544	.594	.667	.760	.833	.901	.967	1.04	1.11	1.21	1.37	1.51	1.64
28	.544	.598	.672	.766	.839	.905	.969	1.04	1.11	1.21	1.36	1.49	1.62
29	.559	.615	.679	.772	.840	.905	.969	1.04	1.11	1.21	1.35	1.49	1.60
30	.559	.615	.682	.774	.845	.908	.971	1.04	1.11	1.21	1.35	1.47	1.59

Table 1. Percentage Points, u_p , such that $\text{Prob}(U_N \leq u_p) = p$

$N \backslash P$	2.5	5.0	10.	20.	30.	40.	50.	60.	70.	80.	90.	95.	97.5
31	.566	.621	.691	.774	.846	.909	.971	1.04	1.11	1.20	1.34	1.47	1.58
32	.574	.625	.691	.780	.847	.911	.972	1.04	1.11	1.20	1.33	1.45	1.57
33	.580	.632	.696	.784	.851	.911	.972	1.03	1.11	1.20	1.33	1.45	1.57
34	.584	.638	.704	.789	.855	.914	.973	1.03	1.11	1.19	1.32	1.44	1.56
35	.586	.642	.707	.790	.856	.914	.975	1.03	1.11	1.19	1.32	1.43	1.53
36	.595	.645	.710	.792	.857	.914	.975	1.03	1.11	1.19	1.32	1.43	1.53
37	.599	.652	.712	.796	.861	.920	.977	1.03	1.10	1.19	1.31	1.42	1.53
38	.599	.656	.717	.800	.863	.920	.978	1.03	1.10	1.19	1.31	1.42	1.51
39	.606	.659	.722	.803	.866	.924	.978	1.03	1.10	1.18	1.30	1.41	1.51
40	.614	.664	.725	.805	.868	.925	.978	1.03	1.10	1.18	1.30	1.40	1.49
41	.615	.670	.729	.807	.869	.925	.978	1.03	1.10	1.18	1.29	1.39	1.49
42	.619	.670	.731	.810	.872	.926	.978	1.03	1.10	1.18	1.29	1.39	1.48
43	.624	.675	.734	.812	.872	.926	.978	1.03	1.10	1.18	1.29	1.39	1.48
44	.624	.675	.737	.816	.874	.927	.978	1.03	1.09	1.17	1.29	1.39	1.48
45	.633	.681	.741	.818	.875	.928	.980	1.03	1.09	1.17	1.28	1.38	1.46
46	.634	.682	.741	.818	.877	.930	.981	1.03	1.09	1.17	1.28	1.38	1.46
47	.636	.685	.745	.819	.878	.930	.981	1.03	1.09	1.17	1.27	1.37	1.46
48	.641	.691	.746	.819	.879	.930	.981	1.03	1.09	1.16	1.27	1.37	1.46
49	.642	.691	.751	.824	.880	.933	.982	1.03	1.09	1.16	1.27	1.36	1.44
50	.642	.694	.753	.827	.883	.934	.983	1.03	1.09	1.16	1.27	1.36	1.44
51	.649	.697	.753	.827	.885	.936	.984	1.03	1.09	1.16	1.26	1.35	1.44
52	.650	.697	.754	.829	.885	.936	.984	1.03	1.09	1.16	1.26	1.35	1.44
53	.652	.699	.759	.831	.886	.936	.984	1.03	1.09	1.16	1.26	1.35	1.44
54	.655	.704	.760	.833	.888	.938	.985	1.03	1.09	1.16	1.26	1.34	1.43
55	.660	.706	.762	.835	.889	.938	.985	1.03	1.09	1.15	1.25	1.34	1.42
56	.659	.706	.764	.837	.890	.939	.985	1.03	1.09	1.15	1.25	1.34	1.42
57	.666	.712	.767	.837	.890	.939	.985	1.03	1.09	1.15	1.25	1.34	1.41
58	.664	.712	.769	.839	.892	.939	.985	1.03	1.09	1.15	1.25	1.33	1.41
59	.668	.716	.771	.839	.893	.939	.985	1.03	1.09	1.15	1.25	1.33	1.41
60	.673	.719	.771	.840	.893	.939	.985	1.03	1.09	1.15	1.25	1.33	1.41

Table 1. Percentage Points, u_p , such that $\text{Prob}(U_N \leq u_p) = p$

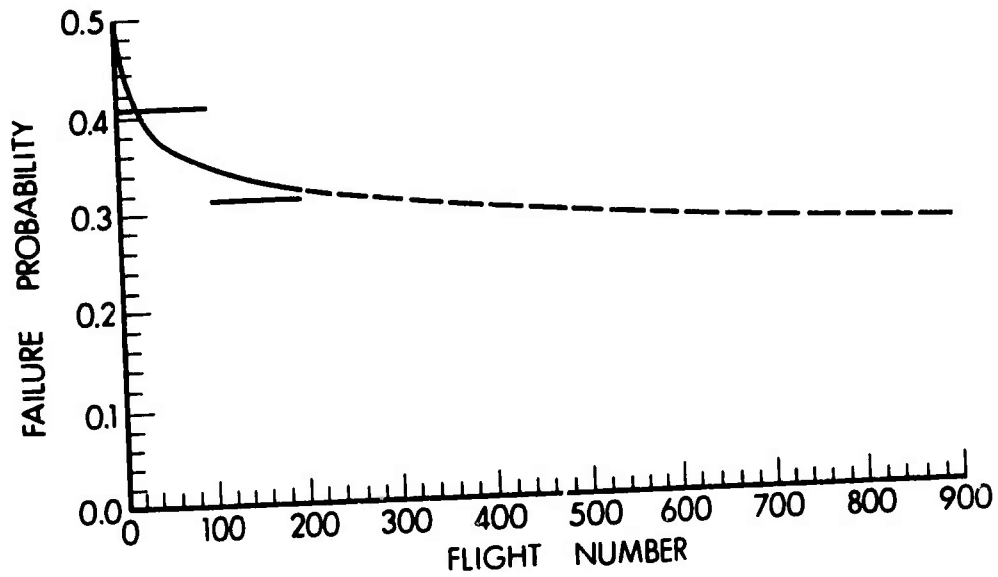


Figure 1. Estimate of Failure Probability for First 200 Valid Flights.

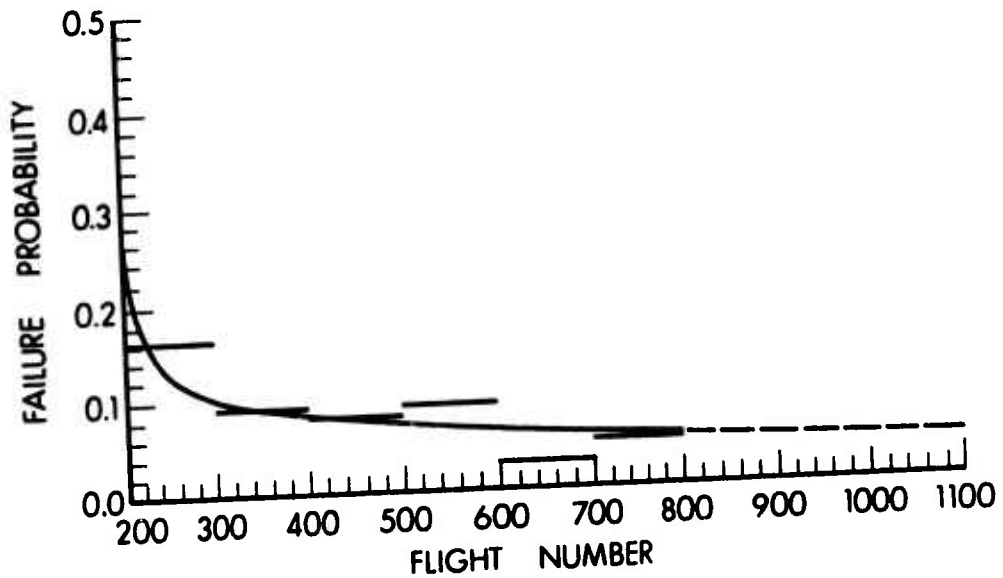


Figure 2. Estimate of Failure Probability Excluding First 200 Valid Flights.

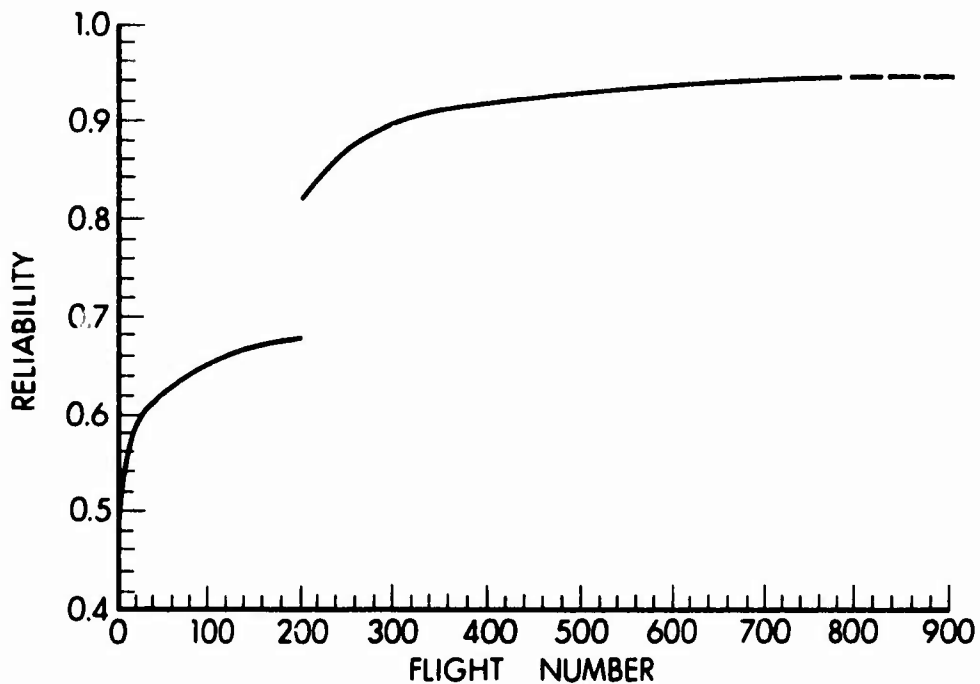


Figure 3. Estimated Reliability Based on Weibull Model.

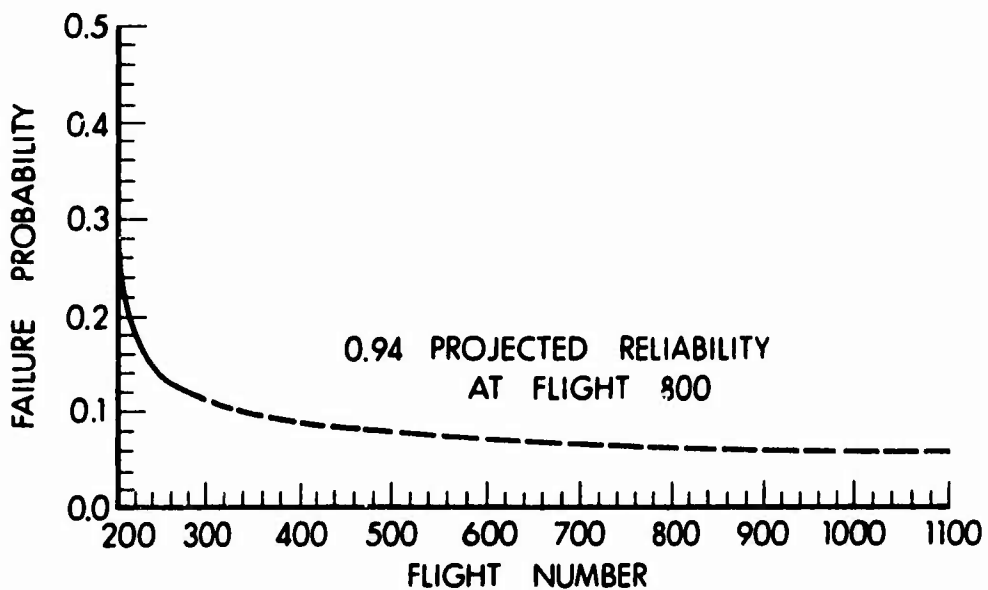


Figure 4. Projected Failure Probability Based on Flights 201 to 300.

IMAGE FORMATION BY MEANS OF SPATIAL INTENSITY CORRELATIONS

PAUL H. DEITZ, PhD
US ARMY BALLISTIC RESEARCH LABORATORIES
ABERDEEN PROVING GROUND, MARYLAND

A large number of the Army's present and future optical systems have as their primary task the image formation of a distant object. In virtually all of these systems, atmospheric turbulence plays a role in limiting the image quality. This is so because the atmospheric medium is inhomogeneous in temperature and, hence, in refractive index. Thus, virtually all military optical systems must operate in media that vary randomly in space and time. Even nonimage-forming optical devices such as laser rangefinders⁽¹⁾ are affected by the grosser manifestations of turbulence which take the form of energy redistribution⁽²⁾ within the optical beam and random pointing of the beam.⁽³⁾ However, for devices such as telescopes that are used to form an image, the random modulation of the optical phase as light traverses the medium acts in a more subtle way to lower the resolving ability of an instrument below its *in vacuo* (diffraction-limited) performance level.

Because of the deleterious effect of turbulence on the performance of image-forming instruments, we have reexamined the techniques of image formation in terms of a new approach to a problem that has afflicted astronomers for over two centuries. We will review the basic mathematical relationships central to classical optical imagery along with the constraints intrinsic to the detection of radiation at optical wavelengths. It will be shown that the only parameter accessible to measure is the intensity (mean square of the electric field) and that this parameter corresponds to a second-order correlation of the electric field.

Having developed the basis of classical imagery, we will introduce the notion of higher-order correlations of optical fields (fourth,

sixth, etc.) and discuss the ways in which these measures relate to standard image formation with which we are familiar. Specifically, by some manipulation of the far-field intensity distribution of a source, we wish to infer the intensity distribution over the source itself. We will discuss some of the constraints of higher-order correlation techniques, the potential they hold for reducing the effects of transmission in the atmosphere, and a basic experiment that illustrates the mathematical relationships.

In this paper we intend to present a mathematical framework upon which we can hang physical and intuitive arguments. Although extensive mathematical treatment will be avoided, annotation has been provided for those who might wish to explore these topics in detail.

1. THE HUYGENS-FRESNEL EQUATION

We begin our discussion of optical imaging by examining the coordinate planes of Fig. 1. The object to be imaged here is found in the ξ - η plane at the left-hand portion of the figure and is bounded by the aperture denoted Σ . In the general case, the object may be a primary source of light or a secondary surface from which light is scattered. The object is described by a two-dimensional source, since an extension along the axis of propagation can be accounted for, here, by an equivalent field at the plane of radiation. It is our interest to describe the nature of the electric field as it propagates from the ξ - η to the x - y plane, a distance, R_0 , away. For the moment, we are concerned only with the form of the electric field $V(x,y)$ as it relates to the field at the object, $V(\xi,\eta)$. The particular length between specific points in the two planes is indicated by the path $R(\xi,\eta;x,y)$.

Upon realizing that each point in the aperture radiates a spherical wave to the right, we can write the Huygens-Fresnel principle⁽⁴⁾ for paraxial waves for which

$$V(x,y;t) = \iint_{-\infty}^{\infty} V(\xi,\eta;t) \frac{\exp[ikR(\xi,\eta;x,y)]}{i\lambda R(\xi,\eta;x,y)} d\xi d\eta, \quad (1)$$

where $k = 2\pi/\lambda$, and λ equals the wavelength of the radiation. Equation (1) states that the electric field $V(x,y;t)$ is formed of a superposition (linear summation) of waves emanating from each point within the aperture, properly phase-shifted according to the exponential term and diluted by the $1/R$ expression. Now

$$R(\xi,\eta;x,y) = [R_0^2 + (x-\xi)^2 + (y-\eta)^2]^{\frac{1}{2}} \quad (2a)$$

$$\approx R_0 \left\{ 1 + \frac{1}{2} \left(\frac{x-\xi}{R_0} \right)^2 + \frac{1}{2} \left(\frac{y-\eta}{R_0} \right)^2 \right\}, \quad (2b)$$

when approximated by the first two terms of the binomial⁽⁵⁾ (the Fresnel approximation) expansion. Using Eq. (2b) in Eq. (1) and making the far-field (Fraunhofer) approximation [$R_0 \gg k(\xi^2 + \eta^2)_{\max}/2$], we can write

$$V(x, y; t) = \frac{\exp[ik(R_0 + (x^2 + y^2)/2R_0)]}{i\lambda R_0} \times \iint_{-\infty}^{\infty} V(\xi, \eta; t) \exp\left[-i \frac{k}{R_0} (x\xi + y\eta)\right] d\xi d\eta. \quad (3)$$

Apart from the coefficients before the integral, Eq. (3) shows that the electric fields in an aperture and the far field are related by a spatial Fourier transform operation. This property is basic to the ensuing work involving intensity correlations with quasimonochromatic light. We note that Eq. (3) is linear; that is, the two-dimensional integral operator on $V(\xi, \eta; t)$ is linear. Therefore, if either $V(\xi, \eta; t)$ or $V(x, y; t)$ is known, the other is specified through a linear (and hence, invertible) transformation.

2. THE VAN CITTERT - ZERNIKE THEOREM

Although Eq. (3) holds for radiation of general frequencies, the relationship it expresses is somewhat academic here from the standpoint that it describes electric-field quantities that are unmeasurable at optical wavelengths. No known detector can follow oscillations at frequencies of 10^{14} Hz. At optical frequencies, the parameter accessible to measurement is the intensity, the mean square of the electric field averaged over many oscillations. It is this constraint that has prompted many investigators to couch optical theory in the form of correlations of field quantities. The best known of these correlations is called the mutual coherence function, $\Gamma(\underline{x}_1, \underline{x}_2, \tau)$, where

$$\Gamma(\underline{x}_1, \underline{x}_2, \tau) \equiv \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T V(\underline{x}_1, t + \tau) V^*(\underline{x}_2, t) dt, \quad (4a)$$

$$= \overline{V(\underline{x}_1, t + \tau) V^*(\underline{x}_2, t)}, \quad (4b)$$

and τ is the time delay between the instantaneous product of the electric fields at the points \underline{x}_1 and \underline{x}_2 (and $\underline{x}_i = x_i, y_i$). A special case

DEITZ

of the mutual coherence function results when the time delay is set to zero or

$$\Gamma(\underline{x}_1, \underline{x}_2, \tau)_{\tau=0} \equiv J(\underline{x}_1, \underline{x}_2), \quad (5)$$

where $J(\underline{x}_1, \underline{x}_2)$ is called the mutual intensity function.

We wish to calculate the mutual intensity in the far field of a spatially incoherent source. Using Eqs. (5), (4b), and (3), we write

$$\overline{V(x_1, y_1; t) V^*(x_2, y_2; t)} = J(x_1, y_1; x_2, y_2) \quad (6a)$$

$$= \frac{\exp\left[\frac{ik}{2R_0} (x_1^2 - x_2^2 + y_1^2 - y_2^2)\right]}{(\bar{\lambda} R_0)^2} \iiint_{-\infty}^{\infty} \overline{V(\xi_1, \eta_1; t) V^*(\xi_2, \eta_2; t)} \\ \times \exp\left[-i \frac{\bar{k}}{R_0} (x_1 \xi_1 - x_2 \xi_2 + y_1 \eta_1 - y_2 \eta_2)\right] d\xi_1 d\xi_2 d\eta_1 d\eta_2, \quad (6b)$$

where the processes of temporal averaging and spatial integration have been interchanged and the radiation assumed quasimonochromatic ($\nu_{\min} \gg \nu_{\max} - \nu_{\min}$) so that the wavelength dependence can be approximated by the mean wavelength. Since the source is spatially incoherent the mutual intensity takes the form⁽⁶⁾

$$\overline{V(\xi_1, \eta_1; t) V^*(\xi_2, \eta_2; t)} = J(\xi_1, \eta_1; \xi_2, \eta_2), \quad (7a)$$

$$= I(\xi, \eta) \delta(\xi_1 - \xi_2) \delta(\eta_1 - \eta_2). \quad (7b)$$

Physically, Eq. (7b) implies that the time fluctuations of the electric fields at two non-identical points in the source plane are completely uncorrelated;* equivalently, the total power measured at a point in the far field is simply the sum of the squared electric fields from each differential element of the source, taken with the proper phase delay and attenuation.

If Eq. (7b) is used in Eq. (6), the mutual intensity collapses to a single area integral, giving

* The overbar of Eq. (7a) by the definition of Eq. (4) indicates a time average. Although spatial incoherence is usually defined by a time-averaging process, it can also be defined in terms of a spatial average in the source plane.

$$J(x_1, y_1; x_2, y_2) = \frac{\exp\left[i \frac{\bar{k}}{2R_0} (x_1^2 - x_2^2 + y_1^2 - y_2^2)\right]}{(\bar{\lambda}R_0)^2} \times \iint_{-\infty}^{\infty} I(\xi, \eta) \exp\left\{-i \frac{\bar{k}}{R_0} [(x_1 - x_2)\xi + (y_1 - y_2)\eta]\right\} d\xi d\eta. \quad (8)$$

Although usually presented in a normalized form,⁽⁷⁾ Eq. (8) is a statement of the van Cittert-Zernike theorem which, aside from the accompanying coefficient, shows that the mutual intensity in the far field of a spatially incoherent source is given by the Fourier transform of the intensity across that source.

The mutual intensity function of Eq. (8) is directly accessible through measurement using a technique known as Young's double pinhole experiment. Figure 1 shows an opaque surface erected in the x-y plane. Two pinholes are made in the surface. A distance behind the surface a fringe pattern can be observed on the u-v plane. The fringe contrast relates to the amplitude of $J(x_1, y_1; x_2, y_2)$, and the fringe shift corresponds to the relative phase between the radiation on the pinholes. If this measurement is made for all pinhole spacings and orientations in the x-y plane, $J(x_1, y_1; x_2, y_2)$ is completely specified and can be Fourier transformed to derive $I(\xi, \eta)$, the intensity on the object. This theorem is basic to most optical imaging, since lenses effect the inverse transformation of Eq. (8) to give a scaled distribution of the object irradiance. Referring again to Fig. 1, we note that a lens has been placed its own focal length from both the x-y and u-v planes. We assert without proof (see Ref. 4, p. 83) that in this configuration there is a Fourier transform relation between these two planes due to the operation of the lens on the radiation from the x-y surface similar to the relationship between the ξ - η and x-y planes located a great distance apart. Thus, under the assumptions given in the mathematical treatment above, the electric field $V(u, v)$ is a scaled distribution of the field $V(\xi, \eta)$. The inversion of the u-v axes is a mathematical convenience (see Ref. 4, p. 167).

By the scaling property from one domain to another intrinsic to the Fourier transform operation, large detail in the source is given by small separations in the far field and vice versa. Thus, imaging with a finite aperture implies a finite limit to the high-frequency detail resolvable on the source. This justifies the well-known description of lenses as low-pass filters. Also, since there are many more pairs of points within an aperture corresponding to small separations than to large, there is a built-in redundancy weighted in favor of low-frequency resolution.

Equation (8), although a statement of second-order correlation, is basic also to fourth-order correlation, as we shall see later.

3. THE EFFECTS OF COHERENCE

At this point in the development, it is important that we concern ourselves with the coherence properties of optical sources.^(8,9) We again refer to Fig. 1. Let us imagine that there is a single point scatterer at the object (ξ - η) plane illuminated by an optical source. The field at the x-y plane (plane of observation) will be, of course, a spherical wave. If we begin to insert arbitrarily other point scatterers at the object plane, the electric field at the x-y plane will be the sum of all the individual contributors as determined by geometry per Eq. (3). At a particular instant of time, the contributions from the various scatterers will add constructively at some points on the x-y plane, while at other points they may not. But, as indicated earlier, the electric field is not accessible to measurement; the intensity (mean square of the electric field) must be averaged over many cycles of field oscillation. If the optical bandwidth of the light illuminating the scatterers is sufficiently narrow that the relative phase across the wavefront remains constant during the period of observation, then the intensity pattern that would be observed during just a few temporal cycles of oscillation will remain fixed during the total period of observation. This pattern is called a speckle pattern. If, on the other hand, the bandwidth of the source is broadened, then the speckle pattern begins to change with time and gradually washes out unless the period of observation is shortened accordingly.

Fluctuations of the speckle pattern in the time domain relate to the temporal coherence properties of the source. Fluctuations of the speckle pattern in space (across the x-y plane) relate to the source size and are termed spatial coherence effects. It is these latter coherence properties that are embodied in the mutual intensity function of Eq. (8) and relate directly to the parameter of interest, $I(\xi)$, the intensity distribution over the source plane. But the ability to record the speckle pattern without its being washed out depends on the source bandwidth and detection time in the x-y plane.

Speckle patterns are easily observed by eye when laser radiation is scattered from a rough surface. However, with thermal sources great care must be taken to utilize even a fraction of the speckle signal that is usually time-variant with extremely short time constants.

4. A THEOREM CONCERNING THE FOURTH-ORDER GAUSSIAN RANDOM PROCESS

We have seen in Eq. (7) the way in which a time average can be used to impose a condition on the correlation of an electric-field pair. No particular assumption was made about the statistics of the field variables. However, it is well known that classical thermal sources exhibit statistical fluctuations (temporally) that are gaussian in nature. Hodara⁽¹⁰⁾ has asserted that lasers with but a few axial modes are, to a good approximation, gaussian as well. However, Troup⁽¹¹⁾ has argued that gaussian statistics are achieved only in the limit of a large number of axial modes.

The argument for gaussian statistics can be extended to the spatial domain as well. It has been argued that⁽¹²⁾ the received field at any point in the far zone (as described above in Section 3) consists of a sum of random-amplitude, random-phase, complex phasors contributed by the elementary scatterers. If the size of the scattering area is large enough to include many point scatterers (or there are enough elementary coherence areas composing the source), the Central Limit Theorem may be used to conclude that the electric field in the detection plane is a gaussian random process in a spatial sense.

A well-known property of gaussian statistics is that all higher-order moments are representable in terms of the first and second.⁽¹³⁾ In particular, the fourth-order correlation of electric fields (second-order correlation of intensities) can be shown to be⁽¹⁴⁾

$$\overline{I_1 I_2} = \overline{I_1} \overline{I_2} + |\overline{r_{12}}|^2, \quad (9)$$

where the defining relation of Eq. (4a) has been used. Equation (9) therefore describes the relationship between intensity correlations and field correlations for a process that is gaussian in the time domain. We note that, in general, the field correlation is a complex quantity, so that only the relationship between the intensity correlation and the modulus of the field correlation is implied. It reveals the underlying principle by which intensity correlations in the far field may be used to infer the accompanying field correlations and hence, through Eq. (8), to gain knowledge of the intensity distribution at the source.

5. ASSUMPTIONS BASIC TO STATISTICAL AVERAGING

We are now in a position to examine the nature of the averaging processes fundamental to a number of optical processing schemes. Figure 2 represents an ensemble of similar experiments. As in Section 3, we consider the intensity distribution in the x -plane to be made over

DEITZ

an observation time short compared with any temporal fluctuations in intensity. The amplitude distribution over the source plane is identical from one sample to the next. The pair of receiver points in the detection plane also remains constant. However, each source plane exhibits a statistically similar, but independent phase mapping across its extent. This insures that the speckle (intensity) patterns observed at the receiver (x) planes remain statistically identical as well. In the language of probability theory, each pair of measurements in the x -plane represents one particular sample in the outcome space of the experiment.

We now turn our attention to the classic relation of imaging, Eq. (8). It must be recognized that there are two fundamental assumptions basic to its development which were not explicitly mentioned. First, the mutual coherence function of Eq. (4) is fundamentally defined in terms of an ensemble average. This is the average that would be calculated if the total outcome space for a given experiment were known for a given space-time set of boundary conditions. In practice, the ensemble average is never measurable because of a lack of access to an unlimited number of experimental configurations operating under identical circumstances. Instead, the ergodic hypothesis is invoked: that is, if one particular experiment is performed under unchanging conditions (stationarity), then the average that is taken in the domain of stationary conditions is *assumed* equivalent to the ensemble (true) average. The working definition of Eq. (4) necessarily assumes that the ensemble average can be replaced by a time average because of a condition of stationarity in the time domain, the domain in which the averaging is made. Equation (8) is then a time average over many instantaneous products of electric field. For this situation, Fig. 2 can represent a time series of field pairs (i, j, \dots, n) which are sampled, multiplied, and averaged.

6. FOURTH-ORDER CORRELATIONS IN THE TIME DOMAIN

Until about two decades ago, essentially all imaging was accomplished by means of second-order correlations taken over a time average long compared with the coherence time of the radiation. About that time, two astronomers, Hanbury Brown and Twiss,^(15,16) were searching for a technique to infer the diameter of stars which would not have the sensitivity to atmospheric turbulence and instrument vibration characteristic of the stellar interferometry of Michelson and Pease.⁽¹⁷⁾ To that end, they were first to suggest the use of fourth-order field correlations. The key to this approach is embodied in Eq. (9).

Although Hanbury Brown and Twiss did not explicitly use the gaussian theorem of Eq. (9), it is, in fact, intrinsic to their

DEITZ

mathematical development. Their result shows, nevertheless, that the time-averaged, two-point correlation of intensities in the far field of a spatially incoherent source is proportional to the square of the Fourier transform of the source intensity distribution. We note as before that Γ_{12} is, in general, a complex function which, in fact, represents the amplitude spectrum of the source intensity distribution. Since in Eq. (9) the mutual coherence function is squared, this method yields the power spectrum of the source intensity distribution. Also, since the relative magnitudes of the real and imaginary components of the amplitude spectrum are unknown, no linear transformation is possible to infer the intensity distribution of the source itself.

Turning again to Fig. 2, the Hanbury Brown-Twiss experiment can be understood by letting each member of the ensemble represent a pair of intensity measurements made in a time less than the coherence time of the radiation. As many intensity products are averaged in time, the function approaches the ensemble average.

Although this method of imagery yields information only about the object intensity power spectrum, it has proved useful in the measurement of star diameters,⁽¹⁸⁾ for which only the first zero crossing of the Fourier transform need be known. Since the optical phase is discarded immediately upon detection, the instrumentation is not only insensitive to mechanical vibrations but to turbulence-induced phase fluctuations as well. The primary limitation to the method is the low level of radiation that is detected. As a result, this approach is limited to relatively intense stars to overcome the severe signal-to-noise problem. Analyses of this problem have been given by Gamo⁽²⁰⁾ and Twiss.⁽²¹⁾

7. INTENSITY CORRELATION IN THE SPATIAL DOMAIN

With the utilization of the laser, the experimental constraints are quite different. The signal-to-noise ratio can be increased typically by six orders of magnitude.⁽²²⁾ With a view to exploiting this property, Deitz and Carlson⁽²³⁾ have recently investigated the potential for intensity correlation in terrestrial imagery. In many situations of interest to the Army, the availability of laser illumination greatly reduces the constraints due to noise limitations. In addition, since ground-to-ground imagery suffers from the greatest deterioration in image quality because of turbulence, it offers the greatest potential for improvement.

Although the details of this work can be found elsewhere,⁽²³⁾ we will again turn to Fig. 2 for the basis of this recent approach. In the case of spatial intensity interferometry, the signal in the x-plane

is detected over a large area, rather than at just two points. Conceptually, a large piece of film is placed in the \underline{x} -plane. The signal is recorded spatially over a time short compared with the coherence time of the radiation and is then autocorrelated. This operation corresponds to an average taken in the space domain. Here, through the assumption of spatial stationarity (and a sufficiently large area of spatial averaging), the spatial average is assumed equivalent to the ensemble average.

Although the method of signal processing is very different from the Hanbury Brown - Twiss experiment, the final mathematical relations are similar. The principal result⁽²³⁾ shows that the spatial intensity correlation in the far field of a source can be written

$$\langle I(\underline{x}_1, T) I(\underline{x}_2, T) \rangle = \frac{1}{(cr)^4} \left| \int_0^\infty \omega^2 d\omega \int_{-\infty}^\infty \text{sinc}(\rho T) H(\rho) d\rho \right. \quad (10a)$$

$$\begin{aligned} & \times \iint_{-\infty}^\infty I(\underline{\xi}, \omega) \exp\left[-i \frac{k}{r} (\underline{x}_1 - \underline{x}_2) \cdot \underline{\xi}\right] d\underline{\xi} \\ & \times \iint_{-\infty}^\infty C(\underline{f}) \exp\left[-i \frac{k}{2r} (\underline{x}_1 + \underline{x}_2) \cdot \underline{f}\right] d\underline{f} \Big|^2 \\ & \propto \left| \hat{I} \left[\frac{k}{r} (\underline{x}_1 - \underline{x}_2) \right] \right|^2 \left| \hat{C} \left[\frac{k}{2r} (\underline{x}_1 + \underline{x}_2) \right] \right|^2. \quad (10b) \end{aligned}$$

The angle brackets indicate a spatial average, c is the vacuum velocity of light, r is the average distance between object and detection planes, and ω is the angular frequency of radiation. The parameter ρ is the difference frequency between modes in the incident radiation, the H function is defined in Ref. 23 and relates to the temporal spectrum of the carrier, and k is the wave number of light ($2\pi/\lambda$). Equation (10) shows that the spatial intensity correlation is proportional to two functions. The first is the square of the Fourier transform of the source intensity distribution as related in Eq. (9) by a temporal averaging process. The second term, $C(\underline{f})$, is a normalized phase correlation function⁽¹²⁾ describing the coherence interval over a rough surface. In order to describe a spatially incoherent surface, $C(\underline{f})$ (which is defined by a spatial average) is often allowed to assume the role of a delta function. In that idealized limit, its transform becomes a constant and does not limit the detectable spatial frequency spectrum of the object.⁽²³⁾

8. AN EXPERIMENT IN SPATIAL INTENSITY INTERFEROMETRY

To illustrate the results of Eq. (10), a pair of crossed Ronchi rulings, shown in Fig. 3, were used as the object in the following

experiment.* First, in order to demonstrate the Fourier transform relation of the Huygens-Fresnel equation [Eq. (3)], the object of Fig. 3 was transilluminated. In the far field, the intensity distribution was recorded and is shown in Fig. 4. This signal corresponds to the square of electric field represented by Eq. (3); equivalently, the signal of Fig. 4 is the power spectrum of the electric field distribution over the object of Fig. 3.

To illustrate the results of intensity correlation, a section of ground glass was inserted (to gain spatial incoherence) adjacent to the object of Fig. 3 and again illuminated by a laser. The far-field speckle pattern that resulted is shown in Fig. 5. The pattern appears random. The mathematical effect of the ground glass has been to multiply the object field, $V(\xi, \eta)$, inside the integral of Eq. (3) by a random phase function.

Next, the speckle pattern of Fig. 5 was used to make an identical pair of optical transparencies. These were then used in an optical autocorrelator to record the result shown in Fig. 6. As described by the relation of Eq. (10), Fig. 6 gives the power spectral density of the intensity distribution over the object. Since the electric field amplitude and the intensity over the object of Fig. 3 are related by a squaring process, the results of Figs. 4 and 6 are similar in the manner indicated by Eqs. (3) and (10).

9. THE INVERSION OF THE POWER SPECTRUM

As we have shown in the previous sections, intensity correlations lead only to the power spectral density of the source intensity or, equivalently, the modulus of the Fourier transform. Depending upon the application, the power spectral density or its Fourier transform, the correlation function, may be the parameter sought for system use.⁽²⁴⁾ This is often the case in a target-guidance or recognition problem. However, without the phase information, the Fourier inversion cannot be taken to derive the source distribution itself.

In the experiments of Hanbury Brown and Twiss, the loss of phase is not a serious limitation since their objective is simply the measurement of star diameters. If a circular disk is used as a model for a star, the object is known, *a priori*, to be symmetrical. Thus the spatial transform of the (real) intensity is pure real. For this situation, the phase of the transform is zero or π for all spatial wave numbers, and the square root of the power spectrum can be taken (with a sign ambiguity) to derive the spatial transform itself.

*The author is indebted to N. A. Peppers, Stanford Research Institute, for a number of helpful suggestions in this experiment.

DEITZ

In the application of intensity interferometric techniques to terrestrial imaging systems, though, the loss of phase is a serious limitation to the method. A number of authors have addressed themselves to the problem of phase recovery.⁽²⁵⁾ A general solution to this problem, however, has never been found.

Deitz and Carlson have recently proposed a scheme⁽²⁶⁾ whereby the incoming electric field is preprocessed before intensity detection and autocorrelation. By this method, the source is effectively symmetrized in its intensity distribution. If a function is real and exhibits even symmetry,⁽²⁷⁾ its Fourier transform must be pure real. Thus, even though the power spectrum is finally derived by the measurement method, its accompanying amplitude spectrum is known (by virtue of the preprocessing) to be pure real. Thus, the square root of the power spectrum can be taken to infer the amplitude spectrum. There is a critical choice of signs, however, to be made in the square root process. The utility of the method appears to rest with the degree to which system noise can be precluded from interfering with those decisions.⁽²⁶⁾

10. CONCLUSIONS

There are a number of special benefits from detecting images by the technique of intensity correlation. (1) The method is relatively insensitive to the effects of atmospheric scintillation.⁽¹⁹⁾ (2) Because the signal is detected in the spatial-transform domain, high-frequency detail about the scattering surface translates to large spatial lags in the far field. This result could be particularly important at frequencies where detector resolution is not well developed. (3) A special advantage to intensity interferometry in the spatial domain is the utilization of gaussian statistics in the spatial (not temporal) sense. By this method, sources with non-gaussian time statistics (such as single-axial-mode lasers) can be utilized. (4) Still another advantage of spatial detection is that images of moving surfaces can be formed using brief exposures.

Against these benefits must be weighed the limitations of intensity interferometry. The primary factor in this respect rests upon the greatly diluted energy density at the plane of detection. Because the signals are of such relatively low intensity, they tend to be masked by noise. The limitation can result from quantum noise in the carrier, detector noise, or stray signals from unwanted background. It remains to be seen whether these limitations can be successfully overcome so that the potential benefits of intensity correlation can ultimately be realized.

REFERENCES

1. For an illustration of turbulence effects on the beam propagation of the Army XM-23 E2 Range Finder, see P. H. Deitz and N. J. Wright, *J. Opt. Soc. Am.* 59, 527 (1969).
2. For an analysis of the effects of turbulence on range finder induced ocular damage, see P. H. Deitz, *Appl. Opt.* 8, 371 (1969).
3. E. C. Alcaraz and P. M. Livingston, Ballistic Research Laboratories Memorandum Report No. 2103, June 1971.
4. J. W. Goodman, *Introduction to Fourier Optics* (McGraw-Hill, New York, 1968), Ch. 4.
5. *Mathematical Tables*, C. D. Hodgman, ed. (Chemical Rubber Publishing Co., Cleveland, 1958), p. 296.
6. L. Mandel and E. Wolf, *Rev. Mod. Phys.* 37, 231 (1965).
7. M. Born and E. Wolf, *Principles of Optics* 4th ed. (Pergamon Press, Oxford and New York, 1970), p. 510.
8. W. Martienssen and E. Spiller, *Am. J. Phys.* 32, 919 (1964).
9. M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, 1964), p. 29.
10. H. Hodara, *Proc. IEEE* 53, 696 (1965).
11. G. J. Troup, *Proc. IEEE* 53, 1732 (1965).
12. J. W. Goodman, *Proc. IEEE* 53, 1688 (1965).
13. D. Middleton, *Introduction to Statistical Communication Theory* (McGraw-Hill, New York, 1960), p. 343.
14. P. H. Deitz, "Intensity Interferometry in the Spatial Domain," Ph.D. Dissertation, University of Washington, December 1973.
15. R. Hanbury Brown and R. Q. Twiss, *Phil. Mag.* 45, 663 (1954).
16. R. Hanbury Brown and R. Q. Twiss, *Proc. Roy. Soc., Ser. A*, 242, 300 (1957), and *Ser. A*, 243, 291 (1957). These papers appear in *Coherence and Fluctuations of Light*, Vols. I and II, L. Mandel and E. Wolf, eds. (Dover, New York, 1970).
17. A. A. Michelson and F. G. Pease, *Ap. J.* 53, 249 (1921).
18. R. Hanbury Brown and R. Q. Twiss, *Nature* 178, 1046 (1956).
19. For a discussion of the insensitivity of a fourth-order process on phase modulation, see M. J. Beran and G. B. Parrent, Jr., *Theory of Partial Coherence* (Prentice-Hall, Englewood Cliffs, 1964), p. 173.
20. H. Gamo, *J. Opt. Soc. Am.* 56, 441 (1966).
21. R. Q. Twiss, *Optica Acta* 16, 423 (1969).
22. J. W. Goodman, *Progress in Optics*, Vol. VIII, E. Wolf, ed. (North Holland, Amsterdam, 1970), p. 21.
23. P. H. Deitz and F. P. Carlson, *J. Opt. Soc. Am.* 63, 274 (1973).
24. See, for example, BRL Memorandum Report No. 1741, J.R. Rapp, 1966.
25. E. Wolf, *Proc. Phys. Soc. (London)* 80, 1269 (1962); P. Roman and A. S. Marathay, *Nuovo Cimento* 30, 1452 (1963); A. Walther, *Opt. Acta* 10, 41 (1963); H. Gamo, *Proc. Symp. on Electromagnetic Theory and Antennas*, Copenhagen, June 1962 (Pergamon, New York, 1963), p. 809; H. Gamo, *J. Appl. Phys.* 34, 875 (1963); C. L. Mehta, *J. Opt. Soc. Am.* 58, 1233 (1968).

DEITZ

26. P. H. Deitz and F. P. Carlson, *J. Opt. Soc. Am.* **64**, 11 (1974).
27. R. Bracewell, *The Fourier Transform and Its Applications* (McGraw-Hill, New York, 1965).

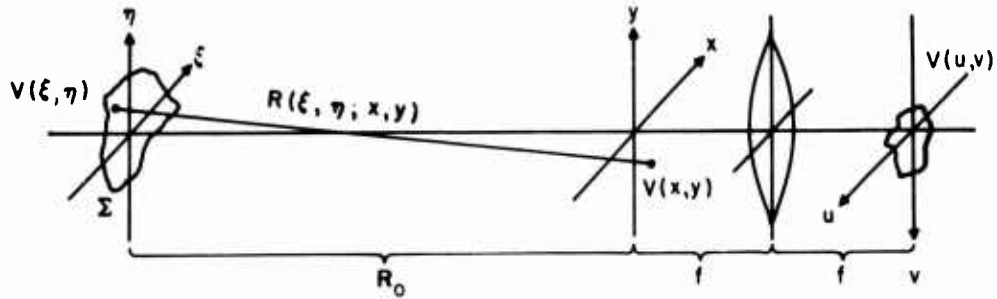


Fig. 1. Coordinate axes for the object (ξ - η) and primary detection (x - y) planes. In addition, a lens is placed its focal length behind the x - y plane. Finally, a focal length behind the lens is the u - v plane.

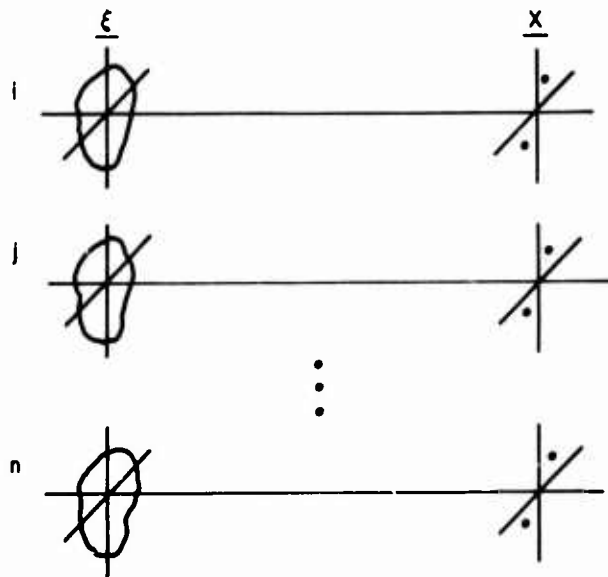


Fig. 2. An ensemble of similar experiments. Source amplitude and receiver points are identical. Random phase variations over each source plane are statistically independent.

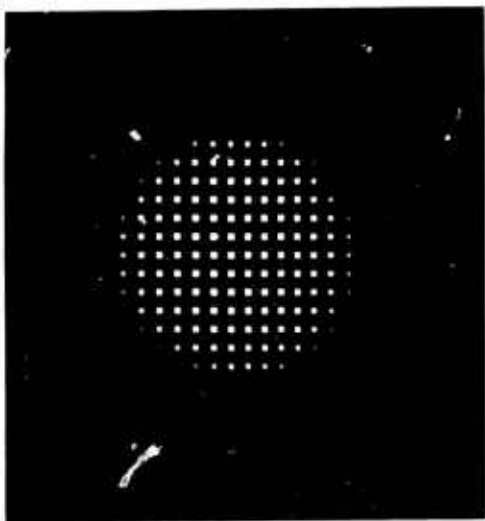


Fig. 3. Optical Transparency
Used as Object

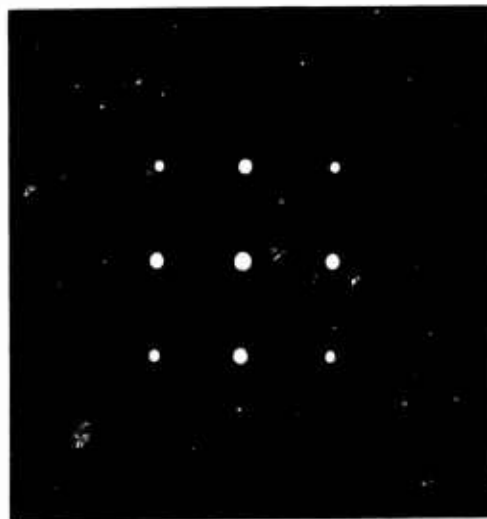


Fig. 4. Spatial Power Spectral
Density of the Electric Field
Distribution of Fig. 3



Fig. 5. Far-Field Speckle Pattern
of Transparency in Fig. 3 When
Used with Random Phase Screen

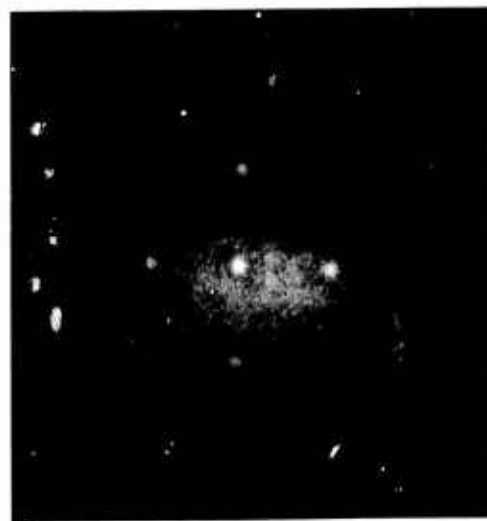


Fig. 6. Spatial Power Spectral
Density of the Intensity Dis-
tribution of Fig. 3

FINITE ELEMENT ANALYSIS OF UNCONFINED FLOW WITH A
VARIABLE TIME-STEP PROCEDURE

C. S. DESAI, Ph.D., U. S. ARMY ENGINEER
WATERWAYS EXPERIMENT STATION, VICKSBURG, MISS.
R. S. SANDHU, Ph.D., OHIO STATE UNIVERSITY
COLUMBUS, OHIO

INTRODUCTION

The phenomenon of flow through porous media is encountered in a number of disciplines of engineering science such as water resources development and hydrology, geotechnical engineering, irrigation and drainage, soil science, hydraulics, oil reservoir engineering, and environmental engineering. The flow problem is often divided into various categories: steady confined and unconfined or free surface; and transient confined and unconfined (1,2,3). The free surface case is governed by nonlinear equations and closed form solutions are difficult or not possible. Such numerical methods as the finite element (FE) and finite difference (FD) procedures are found to be powerful in handling the complexities in terms of material properties, geometrical configurations and nonlinear behavior (1,3).

A FE procedure for the general case of transient unconfined flow is described in this paper; other categories can be handled as special cases of the general case. Numerical characteristics of the proposed procedure are examined and a special scheme is developed that assures satisfaction of convergence and stability at each time level. The aspect of trade-offs in accuracy and computational effort in using approximating models of various orders for the fluid potential is also studied. The validity of the procedure is checked by comparing numerical predictions with laboratory and field observations for typical real life problems.

FINITE ELEMENT METHOD

Governing differential equation

A form of the equation governing the flow in porous (earth) media can be expressed as (1)

$$\frac{\partial}{\partial x} \left(k_x \frac{\partial \phi}{\partial x} \right) + \frac{\partial}{\partial y} \left(k_y \frac{\partial \phi}{\partial y} \right) + \frac{\partial}{\partial z} \left(k_z \frac{\partial \phi}{\partial z} \right) + \bar{Q} = n \frac{\partial \phi}{\partial t} \quad (1)$$

where k_x , k_y , k_z are coefficients of permeability in the x , y , z directions, respectively, n is effective porosity (divided by aquifer thickness) for the unconfined flow, t is time, $\phi = p/\gamma + z$ is fluid head or potential, p is pressure, γ is unit weight, z is elevation head, and \bar{Q} is applied fluid flux. The equation is based on a number of assumptions. Chief among these assumptions are that the medium is incompressible and continuous, fluid is homogeneous and incompressible, flow is continuous and irrotational, capillary and inertia effects are negligible, and Darcy's law is valid. Although the Darcy's law is assumed in the formulation, it is possible to introduce nonlinear or non-Darcy laws in FE formulations (2).

The boundary conditions for transient unconfined flow through an earth dam, Fig. 1, can be stated as

$$\phi = \bar{\phi}_u(t) \text{ on AB} \quad (2a) \quad \frac{\partial \phi}{\partial \eta} = 0 \text{ along free surface CD} \quad (2d)$$

$$\phi = \bar{\phi}_d(t) \text{ on EF} \quad (2b) \quad \phi = z(t) \text{ along CD} \quad (2e)$$

$$\frac{\partial \phi}{\partial z} = 0 \text{ along AF} \quad (2c) \quad \phi = z(t) \text{ along surface of seepage BC and DE} \quad (2f)$$

where η denotes normal to the surface.

Finite element scheme

The finite element method is discussed fully in a number of published works (1). Hence, only a brief description of the main features relevant to the proposed procedure will be covered.

The flow domain is discretized into a number of small subdomains or finite elements. The discretization procedure permits incorporation of values of the unknown ϕ (and its derivatives) at the corners or nodes of the element. The FE method can be formulated for a three-dimensional situation; however, for simplicity, the formulation is presented for two-dimensional flow.

Field variable model. The unknown ϕ within an element, Fig. 2a, is expressed in terms of nodal values of ϕ through a set of interpolation functions (1,4,6-8)

$$\phi(s,t) = \{N\}^T \{q\} \quad (3)$$

where s, t are the natural local coordinates for the element, $\{N\} = [N_1 \ N_2 \ N_3 \ N_4]$ is the vector of interpolation functions N_i , and $\{\phi\}^T = [\phi_1 \ \phi_2 \ \phi_3 \ \dots \ \phi_m]$ is the vector of nodal potentials, and m = number of element nodes. The approximation for ϕ can be generally improved by using higher order functions. The interpolation functions for the three sets of isoparametric elements, Fig. 2a, b, c with 4, 8, and 12 nodes, yielding bilinear, quadratic, and cubic variations of ϕ , respectively, are (1)

$$(1) \text{ 4-node element: } N_i = \frac{1}{4}(1 + ss_i)(1 + tt_i) \quad (4a)$$

$$(2) \text{ 8-node element: } N_i = \frac{1}{4}(1 + ss_i)(1 + tt_i)(ss_i + tt_i - 1),$$

for primary corner nodes

$$N_i = \frac{1}{2}(1 - s^2)(1 + tt_i) \text{ for } s_i = 0 \quad \left. \begin{array}{l} \\ \\ \end{array} \right\} \text{ for secondary (mid-side) nodes}$$

$$= \frac{1}{2}(1 + ss_i)(1 - t^2) \text{ for } t_i = 0$$

(4b)

$$(3) \text{ 12-node element: } N_i = \frac{1}{32}(1 + ss_i)(1 + tt_i) [9(s^2 + t^2) - 10],$$

for primary nodes

$$N_i = \frac{9}{32}(1 - s^2)(1 + 9ss_i)(1 + tt_i), \text{ for } s_i = \pm \frac{1}{3},$$

$t_i = \pm 1$ etc. for secondary $\left(\frac{1}{3}\right)$ point nodes

(4c)

In the isoparametric concept, the coordinates (x, z) of a point in the element are expressed in terms of nodal coordinates by using the same interpolation function N_i ,

$$\begin{Bmatrix} x \\ z \end{Bmatrix} = \begin{bmatrix} \{N\}^T & \{0\} \\ \{0\} & \{N\}^T \end{bmatrix} \begin{Bmatrix} \{x_m\} \\ \{z_m\} \end{Bmatrix} \quad (5)$$

where $\{x_m\}^T = [x_1 \ x_2 \ \dots \ x_m]$ and $\{z_m\}^T = [z_1 \ z_2 \ \dots \ z_m]$.

Element equations

The element equations are derived by using either a variational principle or a residual method such as the Galerkin method (1-3). The transient problem is solved as a series of time independent problems by employing the steady state version of Eq. 1. A variational functional corresponding to this equation is (1)

$$\pi(\phi) = \iint_D \frac{1}{2} \left[k_x \left(\frac{\partial \phi}{\partial x} \right)^2 + k_z \left(\frac{\partial \phi}{\partial z} \right)^2 - 2\bar{q}\phi \right] dx dz - \int_A \bar{q}\phi dA_2 \quad (6)$$

in which D is the flow domain, \bar{q} is the applied flow intensity on a part of the boundary A_2 .

The required gradients of ϕ are obtained by differentiation of Eq. 3 with respect to x, z as

$$\{g\} = \begin{Bmatrix} \frac{\partial \phi}{\partial x} \\ \frac{\partial \phi}{\partial z} \end{Bmatrix} = \begin{bmatrix} [B_1] & [B_2] & [B_3] & [B_4] \end{bmatrix} \{q\} \quad (7a)$$

in which the submatrices $[B_i]$ are given by

$$[B_i]^T = \begin{bmatrix} \frac{\partial N_i}{\partial x} & \frac{\partial N_i}{\partial y} \end{bmatrix} \quad (7b) \quad \text{and} \quad \begin{Bmatrix} \frac{\partial \phi}{\partial x} \\ \frac{\partial \phi}{\partial z} \end{Bmatrix} = [J]^{-1} \begin{Bmatrix} \frac{\partial \phi}{\partial s} \\ \frac{\partial \phi}{\partial t} \end{Bmatrix} \quad (7c)$$

and $[J]$ = the Jacobian matrix.

Substitution of Eq. 7 into Eq. 6 leads to

$$\pi(\phi) = \frac{1}{2} \{q\}^T \iint_D \left(\{q\}^T [B]^T [R] [B] \{q\} - 2 \{q\}^T [N] \{\bar{q}\} \right) dx dz - \int_A \{q\}^T [N]^T \{\bar{q}\} dA_2 \quad (8)$$

Here $[R]$ is the matrix of coefficients of permeability. Minimization of $\pi(\phi)$ with respect to ϕ_i yields element equations as

$$[k]\{q\} = \{Q\} \quad (9)$$

where $[k]$ and $\{Q\}$ are the element property matrix and forcing parameter vector, respectively, and are given by

$$[k] = \iint_D [B]^T [R] [B] dx dz$$

and

$$\{Q\} = \iint_D [N]^T \{\bar{q}\} dx dz + \int_A [N]^T \{\bar{q}\} dA_2$$

The integration is usually performed at local element level with appropriate transformation by using numerical integration. The element equations are assembled and the potential boundary conditions,

Eq. 2a, b, are introduced. The final equations can be expressed as

$$[K]\{r\} = \{R\} \quad (10)$$

where $[K]$ is the assemblage (permeability) matrix, $\{r\}$ is the vector of nodal unknowns for the assemblage and $\{R\}$ is the vector of nodal forcing parameters for the assemblage.

Determination of changing free surface

The free surface experiences movements as a consequence of the variations in the external fluid head. The FE equations, Eq. 10, are solved for ϕ_i ($\{r\}$) and the velocities at the nodes on the free surface, Fig. 3, are computed from (1)

$$\{v\} = -[R]\{g\} \quad (11a)$$

The average Darcy velocities are modified to evaluate the actual particle velocities $\{v_p\}$ as

$$\{v_p\} = \frac{1}{n} \{v\} \quad (11b)$$

The movements of the nodes on the free surface are evaluated by computing displacements normal to the free surface u_n (Fig. 3). The FE mesh is modified or collapsed by computing new coordinates of nodal points on the basis of the displacement components u_r along nodal lines, Fig. 3.

Let the locations of nodes on the free surface be described by a finite dimensional vector, x , consisting of coordinates x_p where $p = 1, 2, \dots, N$, and N is the number of nodes on the free surface. Then by assuming that the changes in geometry are smooth and by using the mean value theorem, the movements of nodes can be expressed as (2,6-9)

$$x_p^{(f)} = x_p^{(i)} + \Delta t \dot{x}_p(t'_p) \quad (12)$$

in which the superscripts (i) and (f) denote the locations of points at time level t and $t + \Delta t$, respectively, Δt is the time increment, $\dot{x}_p(t'_p)$ denotes the rate of change of x_p at time $t'_p \in [t, t + \Delta t]$; \dot{x}_p is determined from the velocities, Eq. 11. A simple method for integration in time would be to compute \dot{x}_p at $t'_p = t$. Such a procedure represents essentially the forward difference integration in time. The numerical stability of this procedure will depend upon the magnitude of Δt , and it will be necessary to preselect a small value of Δt to assure reliable answers. To

avoid the necessity of preselecting values of Δt , an improved scheme called variable time step procedure (10-12) based on the Lipschitz condition and an acceleration method (13) was developed.

Variable time step procedure

A general mathematical form of Eq. 12 can be written as

$$x_p^{(f)} = x_p^{(i)} + \Delta t \dot{x}_p(x_j^{(i)}, x_j^{(f)}, \alpha_p) \quad (13)$$

where α_p are scalar parameters and $x_j^{(i)}$, $x_j^{(f)}$ in the parentheses denote the functional dependence of the rate of change \dot{x}_p on the profiles of the free surface. Since $x^{(f)}$ occurs on both sides of Eq. 13, an iterative algorithm results

$$x_p^{(f)(k)} = x_p^{(i)} + \Delta t \dot{x}_p(x_j^{(i)}, x_j^{(f)(k-1)}, \alpha_p) \quad (14)$$

where k denotes a stage of iterative procedure. The iterations can be started from suitable values of $x_p^{(f)(0)}$ which may be chosen equal to $x_p^{(i)}$.

Convergence of the iterative procedure. Consider two successive iterations, k and $k-1$, expressed first in a general form as (13)

$$x^{(k)} = f(x^{(k-1)}) \quad (15a)$$

and

$$x^{(k+1)} = f(x^{(k)}) \quad (15b)$$

where $x = f(x)$ denotes the iterative solution. For convergence, the following condition should be satisfied (13)

$$\|x^{(k+1)} - x^{(k)}\| \leq \lambda \|x^{(k)} - x^{(k-1)}\| \quad (16)$$

Where λ is the Lipschitz constant which should be less than unity, and $\|\cdot\|$ denotes norm. For the iterative procedure in Eq. 14, the Lipschitz condition can be specialized as

$$\frac{\max_p |x_p^{(f)(k+1)} - x_p^{(f)(k)}|}{\max_p |x_p^{(f)(k)} - x_p^{(f)(k-1)}|} = \Lambda < 1 \quad (17a)$$

in which the norm is defined as

$$\| x^{(f)(k+1)} - x^{(f)(k)} \| = \max_p | x_p^{(f)(k+1)} - x_p^{(f)(k)} | \quad (17b)$$

and $| \cdot |$ denotes absolute value.

Variable time step. In the FE formulation, the values of Λ are computed and compared with unity. If $\Lambda > 1$, the value of Δt is halved such that Eq. 17a is satisfied. If Λ is very large compared to unity, Δt is set equal to $2(\Delta t)$ such that $\Lambda \in [0.5, 1]$. Thus, the magnitudes of Δt are set automatically to assure convergence (and stability), and results into optimization of computational effort.

One of the methods for obtaining \dot{x}_p was to compute it at an average location as

$$\dot{x}_p = \dot{x}_p \left(\frac{x_j^{(i)} + x_j^{(f)}}{2} \right) \quad (18)$$

which when substituted in Eq. 14 yields the iterative scheme as

$$x_p^{(f)(k)} = x_p^{(i)} + \Delta t \dot{x}_p \left(\frac{x_j^{(i)} + x_j^{(f)(k-1)}}{2} \right) \quad (19)$$

Acceleration scheme. The acceleration scheme used in the analysis is based on Aitkin's Δ^2 -method (13), in which the value of α was obtained as

$$\alpha = \frac{x_p^{(k+1)} - x_p^{(k)}}{x_p^{(k-1)} + x_p^{(k+1)} - 2x_p^{(k)}} \quad (20a)$$

which led to

$$x_p = x_p^{(k+1)} - \frac{\left(x_p^{(k+1)} - x_p^{(k)} \right)^2}{x_p^{(k-1)} + x_p^{(k+1)} - 2x_p^{(k)}} \quad (20b)$$

It was found that use of the variable time-step procedure can reduce the number of time steps and can yield equally acceptable accuracy as compared with the simple forward difference procedure, Eq. 12 (10). The basic advantage of the procedure is that one need not preselect a "small" value of Δt .

APPLICATIONS

In order to induce confidence in a numerical procedure, it

is desirable that numerical predictions be compared with laboratory and/or field observations with actual problems. The procedure proposed was used for such correlations for a number of problems (3,5-8, 10,12); only two typical problems are covered briefly in what follows.

Laboratory experiments with parallel plate model

Rise and fall in the river level, as they may occur in the Mississippi River, were simulated in a large viscous flow parallel plate (Hele-Shaw) model, Fig. 4. Detailed description of the model and laboratory tests are given in various reports (5,7). The permeability of the model was evaluated from a specialized form of Navier-Stoke's equations as (5,7)

$$k_m = \frac{b^2 \rho g}{3\mu} \quad (21)$$

where k_m is the permeability of the model, ρ is the fluid density, g is the gravitational constant, μ is the viscosity of the fluid, and b is the half width of gap between the plates. A typical value for k was about 0.236 cm/sec and n was adopted as unity.

The locations of free surface with time were observed and photographed. Typical comparisons between the FE predictions of the free surface and laboratory observations are shown in Fig. 5b. Figure 5a shows the FE mesh used in the analysis.

Field observations at a section along the Mississippi River

A number of sections along the Mississippi River were instrumented to measure fluid heads in piezometers and the corresponding heads in the river. Figures 6a, b show a typical instrumented section with material properties of the layered river bank. The material properties, k and n , were determined from laboratory tests; typical values used were 40 ft/day and 0.4, respectively. The FE procedure was used to predict fluid potentials and location of free surface for a typical history of variation, Fig. 6b, in the river level. In Fig. 7 are shown the FE mesh and comparisons between values of heads and those measured in the piezometers. In both problems, the correlations between predictions and observations are found to be satisfactory.

PARAMETRIC STUDY

A parametric study was performed to examine effects of various factors (12). These included permeability, time increment, rate of drawdown and angle of inclination of the earth bank. A criterion

that a user can employ to arrive at adequate spatial and temporal mesh layouts was derived from these quantitative exercises to be

$$\bar{k}\Delta t \left(\frac{1}{(\Delta \bar{x})^2} + \frac{1}{(\Delta \bar{y})^2} \right) R \leq 0.05 \quad (22)$$

where $\bar{k} = (k_x^2 + k_y^2)^{1/2}$, $\Delta \bar{x}$, $\Delta \bar{y}$ are mean dimensions of the elements, and R is the rate of drawdown.

The problem of trade-offs in accuracy and computational efforts in using approximation models of higher order was examined by using three different isoparametric elements, Fig. 2. It was found that insofar as the accuracy was concerned, all the three elements yielded acceptable solution with no significant improvement in accuracy with higher order models. One of the reasons may be that although the higher orders improve distribution of ϕ within the element and interelement compatibility for ϕ , there seems to be no significant improvement in the computed values of velocities that are based on gradients of ϕ . On the other hand, as shown in Fig. 8, the computational time increased significantly with the order of approximation. For equal number of nodes, the ratios of time taken for the 4-, 8-, and 12-node element was about 1:6:10, whereas, if the number of elements were held constant, the corresponding ratios were 1:15:70. Hence, if a relatively fine mesh is used, the 4-node element can yield acceptable accuracy and economy for the class of problems considered herein with the proposed procedure(12).

CONCLUSIONS

A FE procedure was developed for solution of transient free surface flow through porous media. Such significant characteristics as convergence and stability were considered. In the context of the FE method, a rather new technique based on the Lipschitz condition and an acceleration scheme was introduced. This permitted automatic adjustment of the size of the time increment such that convergence and stability were assured at each time step. The FE procedure was verified by comparing numerical predictions with observations from laboratory and field tests. A parametric study was performed in order to assist the user in understanding trade-offs in accuracy and computational time for different orders of approximations, and in choosing suitable mesh layouts.

Four main aspects of a reliable solution method-formulation, mathematical properties, verification by comparisons with actual problems and applications toward design analysis-were considered. It is believed that the proposed procedure can provide a powerful solution

scheme for the problem of flow through porous media considered in this paper.

ACKNOWLEDGEMENT

The work described herein was sponsored by the U. S. Army Lower Mississippi Valley Division, Vicksburg, Miss., and was conducted by the U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss.

REFERENCES

1. Desai, C. S. and Abel, J. F., Introduction to the Finite Element Method, Van Nostrand Reinhold Co., New York, 1972.
2. Desai, C. S., "Finite Element Methods for Flow in Porous Media," International Symposium on Finite Element Methods in Flow Problems, University of Wales, A Feature State-of-the-Art Paper, Swansea, U. K., Jan. 1974, John Wiley & Sons, London (to appear).
3. Desai, C. S., "Seepage in Geologic Media," in Numerical Methods in Geotechnical Engineering, Desai, C. S. and Christian, J. T. (Editors), McGraw-Hill Book Co., New York (to appear).
4. Zienkiewicz, O. C., Irons, B. M., Ergatoudis, J., Ahmed, S. and Scott, F. C., "Isoparametric and Associated Element Families for Two- and Three-Dimensional Analysis," in Finite Element Methods in Stress Analysis, Holland, I. and Bell, K. (Editors), Technical University of Norway, Trondheim, 1969.
5. Desai, C. S., "Seepage in Mississippi River Banks: Analysis of Transient Seepage Using Viscous Flow Model and Numerical Methods," Miscellaneous Paper S-70-3, U. S. Army Engineer Waterways Experiment Station Vicksburg, Miss., 1970.
6. Desai, C. S., "Seepage Analysis of Earth Banks Under Drawdown," Journal of Soil Mechanics and Foundations Division, ASCE, Vol 98, No. SM11, 1972.
7. Desai, C. S., "Seepage in Mississippi River Banks; Analysis of Transient Seepage Using a Viscous-Flow Model and the Finite Difference and Finite Element Methods," Report 1, Technical Report S-73-5, U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss., 1973.
8. Desai, C. S., "Finite Element Procedures for Seepage Analysis Using an Isoparametric Element," Proceedings, Symposium on Appl. of the Finite Element Method in Geotechnical Engineering, U. S.

DESAI and SANDHU

Army Engineer Waterways Experiment Station, Vicksburg, Miss., 1972.

9. France, P. W. et al., "Numerical Analysis of Free Surface Seepage Problems," Journal of the Irrigation and Drainage Division, ASCE, Vol. 97, No. IRI, 1971.
10. Sandhu, R. S., Rai, I. S., and Desai, C. S., "Variable Time Step Analysis of Unconfined Seepage," Proceedings, International Symposium on Finite Element Methods in Flow Problems, University of Wales, Swansea, U. K., Jan. 1974.
11. Sandhu, R. S., "Analysis of Transient Unconfined Seepage Using Isoparametric Elements (User's Manual)," Report to U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss., 1973.
12. Desai, C. S., Jeng, Y. S., and Sandhu, R. S., "Seepage in Mississippi River Banks; Analysis and Design by Numerical Procedures and Computer Codes," Report 2, U. S. Army Engineer Waterways Experiment Station, Vicksburg, Miss. (in preparation).
13. Henrici, P., Elements of Numerical Analysis, John Wiley & Sons, New York, 1964.

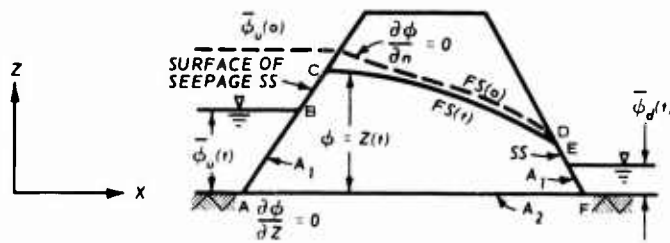
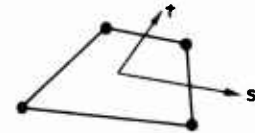
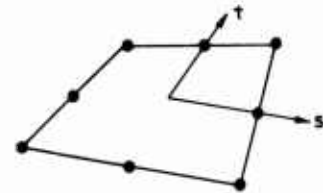


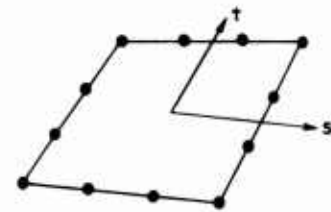
Fig. 1. Boundary conditions for transient free surface flow



a. 4-NODE



b. 8-NODE



c. 12-NODE

s, t = LOCAL COORDINATES

Fig. 2. Isoparametric quadrilateral elements

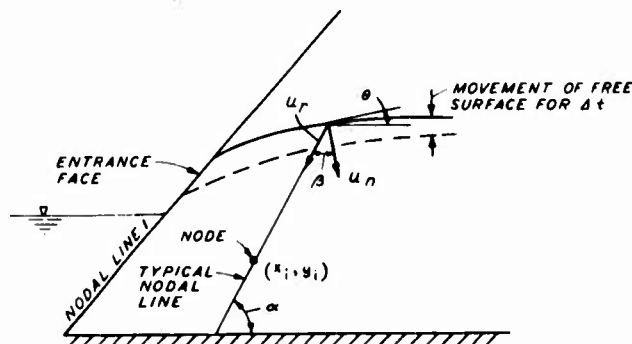
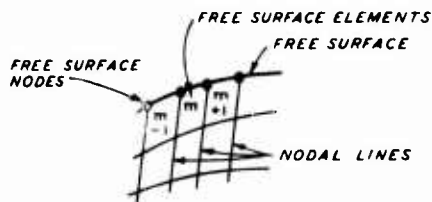
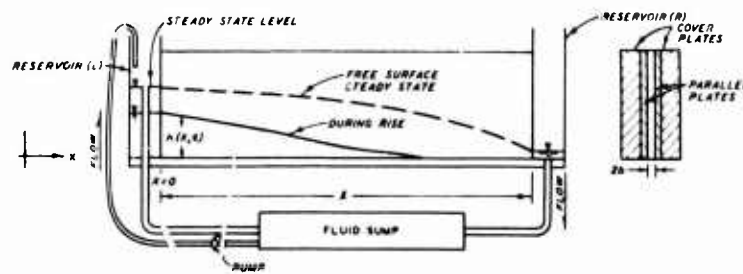
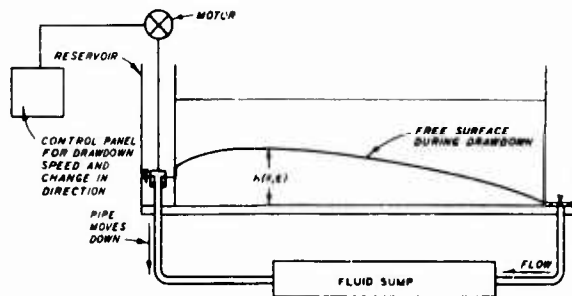


Fig. 3. Movements of free surface

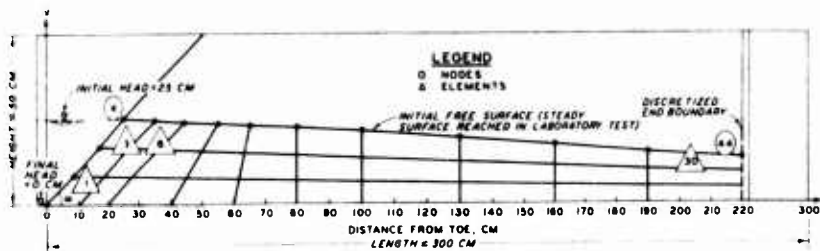


a. RISE IN EXTERNAL FLUID LEVEL

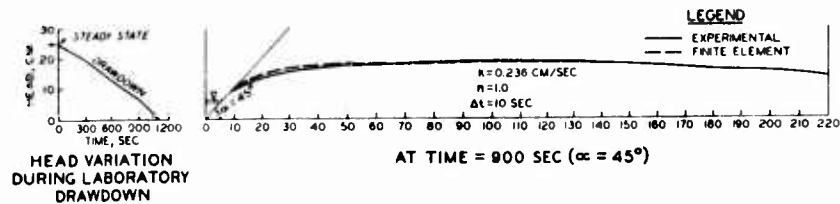


b. DRAWDOWN IN EXTERNAL FLUID LEVEL

Fig. 4. Schematic diagram of parallel plate (Hele-Shaw) model



a. FINITE ELEMENT MESH



b. COMPARISONS BETWEEN PREDICTIONS AND OBSERVATIONS

Fig. 5. Typical comparisons for parallel plate model

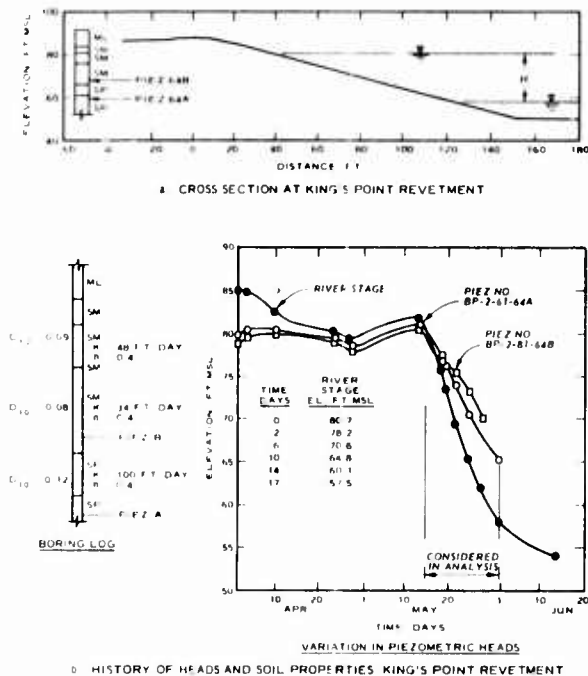
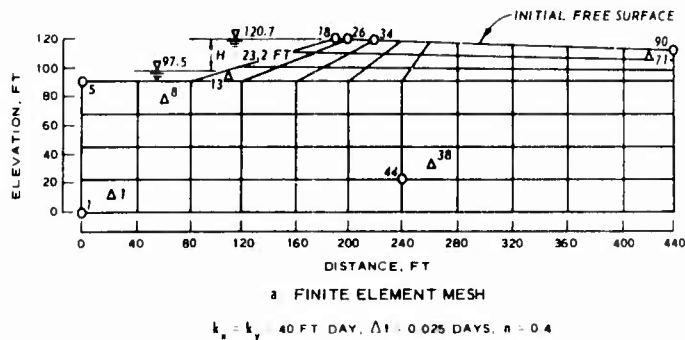


Fig. 6. Details of section and material properties



$k_x = k_y = 40 \text{ FT DAY}$, $\Delta t = 0.025 \text{ DAYS}$, $n = 0.4$

TIME DURING DRAWDOWN DAYS	PIEZOMETER A		PIEZOMETER B	
	MEASURED FT	PREDICTED FT	MEASURED FT	PREDICTED FT
0	80.7	79.0	79.5	79.3
2	79.0	76.9	77.8	77.1
6	74.7	72.3	75.2	72.4
10	70.2	68.8		
14	66.9	66.1		
17	65.0	64.6		

b. COMPARISON BETWEEN PREDICTIONS AND FIELD DATA

Fig. 7. Comparisons for field section

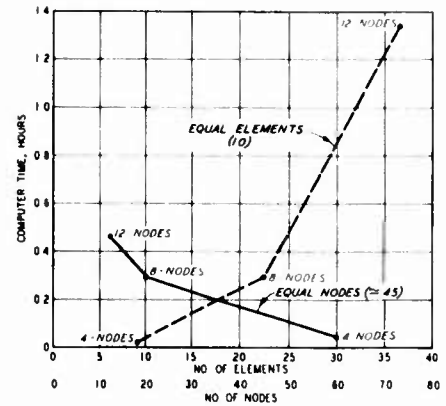


Fig. 8. Comparisons for computer times for three elements

A METHOD OF CONTROLLING SHOCK INDUCED
DAMAGE IN ALUMINUM ALLOYS

ANDREW M. DIETRICH JR., PhD
VICTOR A. GREENHUT, PhD
STANLEY K. GOLASKI
BALLISTIC RESEARCH LABORATORIES
ABERDEEN PROVING GROUND, MARYLAND

In the design of armor to protect a vehicle, suppression of behind the armor fragments (spall resistance) can have significant influence on mission survivability [1]. These fragments increase the probabilities of incapacitating the crew, initiating stowed ammunition, and/or igniting fuel. Although a variety of approaches such as additional ballistic shielding, the use of liner materials or the compartmentalization of fuel and ammunition can be considered to mitigate the effects of spall fragments, much remains to be understood in relating basic armor material properties to spall resistance.

It will be shown that, once a microscopic mechanism is postulated for the occurrence of spall fracture, it is possible to tailor the response of a material to suppress the spall fracture process. Further, it will be demonstrated that the concepts introduced in designing a material are not simply research peculiarities but may be used in a building block fashion to improve material performance across a broad spectrum by tuning the microstructure to react in a desired manner to ballistic loading. Such an approach represents a significant departure from other approaches where conventional material properties like hardness, yield strength, tensile strength, ductility etc. are used as correlating variables to ballistic response. To answer the question of how to inhibit spall fracture, one must first outline a mechanism of its occurrence.

The microstructural damage induced in a metal or alloy during shock loading involves several fundamental mechanisms of flow, and while these are not necessarily unique to shock deformation, the resulting substructures are quite distinct from those arising from

cold work [2,3]. The formation of mechanical twins and dislocation cell structures of small misorientation, as well as the occurrence of diffusionless phase transformations are all documented examples of fundamental flow mechanisms by which metals react to shocks. Another common structural feature is porosity, which is central to the remainder of this paper.

Numerous investigators [4,5,6] have observed the formation of a distribution of pores in a variety of sizes in the region of the spall fracture. These pores have been shown to occur in the zone of interaction between an unloading wave traveling behind the initial compressive shock and a tensile shock reflected from a free (unstressed) surface. Further, it has been shown possible to treat pore formation as a process of nucleation and growth (NAG) in response to the tensile stress-time history. Porosity damage accumulates until a sufficient pore density is present to cause pore link up and the formation of a fracture surface. If porosity damage can be suppressed by restricting its nucleation and/or growth, the tendency of a material to fracture under a given shock condition can be reduced. The question then arises: how can material microstructures be selected to suppress porosity damage?

POROSITY AND FAILURE

Typically the nucleation and growth of pores is characterized by a viscous drag model. One interpretation of the viscous drag model is based on plastic flow resulting from dislocation motion and interaction to form pores. While such a view is quite credible, an alternative microstructural interpretation of the success of nucleation and growth relations can be made.

The authors and co-workers [2,3] during observations of the morphology of shock loading copper single crystals have noted three interrelated microstructural features: an extremely high dislocation density, arranged in a fine cellular structure; local regions of extensive microtwinning; and distributed fine micropores. One striking feature of the micropores in some cubic crystals is their octahedral shape, with facets corresponding to (111) crystal planes. A similar observation has been made by Stevens and co-workers in examining shock-deformed aluminum [6]. They explain their observation with a dislocation interaction mechanism. An alternative view proposed here is that such crystallographic facets represent pore growth faces resulting from vacancy accumulation on (111) planes. Such a mechanism has been shown to cause vacancy loop and/or tetrahedral faulting of rapidly quenched metals. It remains to be shown that an extremely high density of vacancies is produced and that

they migrate to form voids during shock deformation.

The intersection of mobile dislocations leads to the formation of vacancies via jog formation. The extreme density of tangled dislocations, observed as a major component of shock loaded structure, is thus consonant with the formation of a large number of vacancies during passage of an initial compressive pulse. Stress-aided diffusion of vacancies may be expected to enhance vacancy migration rates dramatically at the stress levels characteristic of shock loading. The high tensile stresses resulting from the intersection of the reflected tensile wave and the unloading wave should favor the accumulation of vacancies; and, as previously indicated, the pores may well adopt crystallographic growth facets.

Many metallurgical phenomena such as crystal grain growth, precipitate growth, and stacking fault formation are described theoretically by nucleation and growth laws dependent upon the energies of formation and migration of vacancies. Such vacancy-controlled phenomena usually involve elevated temperature, considerable times and low stresses when contrasted with shock loading times and stresses. An "enhanced" diffusion rate might well be expected at the high stresses and temperatures characteristic of shock deformation. It is thus postulated that pore nucleation and growth in the case of dynamic fracture can be attributed to a stress assisted vacancy diffusion mechanism.

In order to investigate this hypothesis, it was decided that a suitable test required the introduction of a variable factor in a metallurgical system to control the diffusion of vacancies and their accumulation at local sites to form pores. Ideally such a factor should:

1. function as an internal marker of vacancy activity;
2. distribute vacancies as uniformly as possible in the metal;
3. act as a potential sink for vacancies;
4. compete with pore formation and growth by occupying vacancies in a "competitive" diffusion;
5. vary its effectiveness in a controlled manner;
6. demonstrate control of pore formation, growth, joining and resulting spall fracture.

While several possible microstructural and macrostructural systems suggested themselves, a most suitable one could be invoked by applying developments in the field of nuclear technology [7].

Irradiation of metals may lead to the displacement of atoms resulting in the formation of vacancies and interstitials. Such point defects diffuse in the metal to form undesirable voids. It has been found that the introduction of fine, uniformly distributed, metastable precipitates which grow by the mechanism of vacancy diffusion and act as vacancy sinks are an effective means of controlling void formation [7,8,9].

A metal system containing fine, metastable precipitates was examined in order to determine whether such precipitates would function effectively as "vacancy getters" during shock loading and thereby control porosity and spall fracture. They were also employed as internal markers to examine the function of a vacancy mechanism in dynamic loading. Moreover, because they model many commercial alloys they indicate possible immediate applications of the vacancy getter concept to control shock failure. The results and conclusions of these experiments follow.

EXPERIMENTAL PROCEDURE

Al-4 1/2 wt.% Cu was selected as the material for testing. By selection of suitable heat treatments prior to shock-loading a variety of homogeneous precipitate distributions with differing potential as vacancy getters may be obtained. The alloy has been the subject of extensive investigation [10,11]. The precipitate structures and static properties have been well characterized. Moreover, it is a simple binary alloy model for many important commercial aluminum alloys and is typical of a wide range of other aluminum and non-ferrous alloys. The alloy shows a precipitate structure which varies from metastable copper-rich G.P. zones, coherent θ'' , semi-coherent θ' to stable incoherent $\theta(\text{Al}_2\text{Cu})$ upon heat treatment. The precipitates grow at elevated and room temperatures to the various metastable forms dependent on the diffusion of copper. Moreover, the precipitates are thought to act as vacancy traps due to favorable solute-vacancy interaction.

The alloy, fabricated as Instron tensile specimens and shock loading specimens, was heat treated to three conditions designated under-, critical- and over-aged. The alloy was solution treated at 525°C and step quenched (170°C/1 min) to room temperature in order to homogenize the precipitate structure and minimize the precipitate size. The alloy was then held at room temperature for more than 48 hours. This resulted in a homogeneous G.P. precipitate structure with a tensile yield stress of 1.31 MPa (19 ksi) corresponding to the under-aged condition. The alloy was then aged 170°C/8 hr. to

grow θ'' precipitates, i.e. critical-aged, with a maximum yield strength of 1.86 MPa (27 ksi) and finally 300°C/4 1/4 hr. to grow θ' precipitates, over-aged, with a yield stress equivalent to the under-aged condition.

SHOCK LOADING EXPERIMENT

The general experimental configuration is that described in reference [4] and consists of an explosive plane wave lens (TNT, COMP-B) a cylindrical buffer block to attenuate the explosive shock, and a cylindrical sample surround block with the 25.4mm long x 12.7mm diameter sample inserted in the surround block. All interfaces are hand polished. The compressive shock wave generated at the explosive/buffer block interface is reflected from the free surface as a tensile shock wave. The samples are recovered in water. An initial set of experiments was run to determine at what buffer block thickness the over-aged sample would exhibit complete fracture. This thickness was determined to be 85.7mm (3.375 in). The under-aged and critical-aged specimens were then tested for this same set of conditions in order to provide a rapid, preliminary evaluation of the efficacy of the proposed damage suppression concept. As described in subsequent sections of this paper, the anticipated damage levels can be ranked based on initial alloy conditions so that by contrasting the response of the remaining two conditions to the response of the over-aged condition a critical comparison could be made. It is recommended that future tests be conducted with loading geometries more amenable to one-dimensional analysis so that microstructural parameters can be related to damage nucleation and growth through a better documented stress-time history. The use of lower impulses than required for complete spall fracture should be employed to establish early damage nucleation and growth effects.

POST LOADING ANALYSIS

Subsequent to shock loading, the macroscopic and submicroscopic morphology of the specimens was characterized employing a combination of optical methods and electron optical methods. Transmission electron microscope (T.E.M.) observations were made employing a JEM 200KV electron microscope. The results of observation after loading were contrasted with observations of undeformed material, and comparisons were made of the shock loading response of the three precipitate distributions: under-aged (G.P. zone), critical-aged (θ'') and over-aged (θ').

A series of results was anticipated based upon the vacancy getter concept. First, it was expected that if an extremely high rate of

generation of vacancies and stress-aided vacancy migration existed, there should be a significant growth of precipitates relative to the undeformed material. If this precipitate growth characterizes an effective vacancy getter, then porosity should be most inhibited in the case of the under-aged material and porosity should be maximum in the over-aged condition. This phenomenon may be understood to result because the energy which drives precipitate growth arises from the reduction of the surface free energy relative to the volume free energy of precipitates. The finely dispersed G.P. zones (under-aged sample) should thus prove most effective in interfering with the vacancy migration to pores. Further, it might be expected that the inhibition of pore nucleation and growth should lead to the greatest suppression of porosity damage for the under-aged alloy in spite of the fact that the greater tensile yield strength of the critical-aged and higher work-hardening and ultimate strength of the critical- and over-aged specimens might imply more resistance to plastic deformation and failure.

RESULTS

Electron microscopy of the three alloy configurations prior to shock loading reveals that the under-, critical- and over-aged conditions correspond to homogeneously dispersed G.P. zone, θ'' , and θ' precipitates respectively. Because the under-aged, G.P. zone structure may be expected to yield the maximum relative change in precipitate size distribution, attention was concentrated on characterizing any observable change subsequent to shock loading.

T.E.M. micrographs were obtained from undeformed and shock-loaded under-aged material (in the vicinity of the spall fracture region). The observations were made with a similar $\bar{g}=[200]$ two-beam case with beam (zone) axis close to $[001]$ for both samples. Typical bright field (B.F.) and dark field (D.F.) micrographs are shown in Figure 1. It may be noted from the B.F. and D.F. pairs that significantly larger precipitates can be distinguished subsequent to shock loading. In the shocked samples a high density of dislocations was also noted and many precipitates displayed dislocation loops surrounding the precipitates. The micrographs of Figure 1 were taken with conditions to maximize precipitate contrast, and thus those dislocation features are not readily apparent in the micrographs shown. The dislocation loops may be consonant with vacancy accumulation at the precipitate-matrix interface although the possibility of an unexpected dislocation-precipitate interaction under shock conditions cannot be ruled out. Attention should also be drawn to area P of the shock-loaded micrograph (B.F.). This light area may be noted to consist of several regular geometrical shapes. These areas remain bright when the foil

is tilted out of the diffracting condition. The size and traces of the areas at P together with the fact that their contrast must be due to absorption contrast lead to the conclusion that these are voids (or micropores) many of which are included in the foil. Many such areas are noted and indicate that pores grow by void coalescence. These observations are consistent with a vacancy mechanism of pore growth.

In order to characterize the changes in precipitate size and distribution critically, counts were made of the maximum dimension of the disc-shaped particles. A typical precipitate size-number distribution normalized to the total number of counts is shown in Figure 2. The undeformed, under-aged alloy shows a disk diameter maximum population at 80Å which corresponds to a zone thickness of less than 6Å [10]. Thus, the volume of this G.P. zone is less than $3 \times 10^4 \text{ Å}^3$. During shock loading θ'' precipitates are formed and a θ'' precipitate maximum population was found at 280Å which corresponds to a zone thickness of approximately 15Å. The volume of this θ'' precipitate is thus approximately 10^6 Å^3 . The maximum population particle has increased in size by more than 30 times!

The growth in the precipitate size during shock deformation must occur by a process of ripening. Thus, the steep rise in particle number at lower values of precipitate size is due to the fact that some precipitates grow at the expense of others (i.e. ripening). Since the growth of precipitates occurs by solute transport to the larger particles, vacancies must be occupied in compensatory reverse motion. This observation confirms the fact that large numbers of vacancies are created in shock loading with rapid migration rates. The vacancies may be expected to be generated by dislocation intersection, and vacancy-interstitial reaction; the rapid migration is probably stress aided.

These results demonstrate that the precipitates do act as internal markers of vacancy activity, distribute vacancies in the metal, and may act as a potential sink for vacancies (e.g. dislocation loops observed around precipitates). Moreover, it could be anticipated that the precipitates act as vacancy getters in competition with pores.

In order to evaluate the effectiveness of precipitates in controlling porosity damage, metallographic cross sections of the shock loaded samples were evaluated. These observations are shown by Figures

3 and 4. It may be noted that the porosity is least extensive in the under-aged specimen if an examination of the longitudinal extent of observable porosity above and below the spall zone is made. The over-aged specimen shows the maximum degree of porosity. These results indicate that the under-aged precipitate structure was most effective in controlling pore growth, while the over-aged structure was least effective. This is in agreement with the anticipated result that the most finely dispersed precipitates should act most efficiently as vacancy getters.

Relative necking was also measured as an indication of the degree of internal porosity generation and shock damage resistance. To some extent these results were obscured by the amount of plastic flow in late stages of fracture. The observations are summarized in Figure 5. It may be noted that the under-aged material showed less necking than the over-aged. This compares well with the porosity observations. However, the critical-aged alloy shows less necking than either of the other specimens.

This can be understood by examining the specimen cross-sections of Figure 3. The under-aged specimen has a major crack which has opened ductilly. This leads to more extensive necking compared to the critical-aged material in spite of the fact that porosity is lower in the under-aged specimen.

Figure 3 also shows the extent of failure control. The over-aged alloy has failed completely and shows further failure cracks which have resulted from the joining of pores. The critical-aged specimen shows almost complete propagation of spall failure across the specimen and indeed the crack has propagated to the surface at least halfway around the specimen. The under-aged specimen while showing opening of a major spall crack has considerable material integrity to the specimen surface. Thus, as anticipated the model of the precipitate as an effective vacancy getter and consequent failure control parameter appears viable. Design of shock resistant materials and material treatments to take advantage of vacancy gettering and porosity control appears to show an important potential.

SUMMARY AND RECOMMENDATIONS

It has been shown that the microstructure of a metal can be manipulated in a rational, systematic manner to control the level of damage induced by a shock wave. This control has been achieved by first postulating the mechanism by which damage is induced [in this

case porosity damage modeled as resulting from vacancy coalescence during shock wave intersection] and then determining a microstructure [controlled precipitates] to compete for the vacancies and thus suppress damage. The particular alloy studied was selected to provide an uncomplicated model of the proposed process while retaining a major feature of commercial alloys. The experiments to-date indicate that:

- a. Vacancies are significant to the nucleation of porosity damage during shock loading as evidenced by strong precipitate growth in the under-aged alloy.
- b. At the shock strengths examined the potential for vacancy gettering action is more important to suppressing damage than are yield strength, ultimate strength, work hardening rate or ductility as shown by comparison of the results of the under-, over-, and critical-aged alloys.

This initial study has opened the door to a new concept in tailoring material response to ballistic loading. Two avenues can be fruitfully pursued. First, a detailed analysis of shock induced damage in vacancy gettering materials should be conducted to relate the control mechanism to the nucleation and growth parameters. Such a relation would assist in the interpretation of nucleation and growth sensitivity to fundamental material properties and perhaps, stimulate new ideas for control techniques. Secondly, while the alloy selected produces precipitates characteristic of those in the 2000 series of aluminum alloys, many other possibilities exist. Of special interest are the aluminum-zinc and aluminum-silicon binaries and the aluminum-zinc-silicon ternary. These alloys can also be used to model fundamental reactions occurring in commercial materials. Further, the aluminum-zinc-silicon system may be heat treated to produce both a vacancy getter phase (Al-Zn precipitate with maximum coherency analogous to the Al-Cu phase discussed above) and a dispersion hardening phase (incoherent Si particles finely dispersed to increase material shear strength but too small to act as crack nuclei). Since the relative particle sizes and distributions of these two precipitate phases can be controlled independently over a broad range, this alloy's structure can be manipulated to provide an optimum vacancy gettering action for spall fracture resistance without necessarily trading off shear strength.

It is hoped that, in addition to the possible applications of specifically developed alloys, a new approach can be established with respect to materials design and selection for ballistic loading: an approach based on tailoring material structure for a desired ballistic response.

ACKNOWLEDGMENTS

The authors wish to thank Drs. S. Weissmann and C. M. Glass for stimulating our approach to this problem. The concept of "vacancy gettering" as a ballistic material property was derived in the course of many helpful discussions with them on the subject of shock loading response of metals. Our thanks are also extended to Ms. N. Steiner for typing the manuscript.

REFERENCES

1. Zeller, G., "Effect on Tank Vulnerability of Suppressing Behind Armor Spall Fragments (U)" BRL Memo Report No. 1426, Conf. 1962.
2. Kingman, P. W., Metallurgical Effects at High Strain Rates (1972) pp 659 - 668.
3. Dietrich, A. M., Greenhut, V. A., Ibid pp. 645 - 658.
4. Glass, C. M., Moss, G. L., Misesy, J. J., and Golaski, S. K., Dynamic Behavior of Materials, ASTM Publication (1963) pp 282-305.
5. Shockey, D. A., Seaman, L., and Curran, D. R., Metallurgical Effects at High Strain Rates (1972) pp 473 - 499.
6. Stevens, A. L., Davison, L., and Warren, W. E., Jour. App. Phys. 42 p 4922 (1972).
7. Weissmann, S., Private Communication.
8. Norris, D. I. R., Radiation Effects, 14 p 1-37, 15, 1-22 (1972).
9. Sturcken, E. F., Krapp, C. W., and Alewine, G. B., Nuclear Metallurgy, 19 pp 108 - 132 (1973).
10. Kelly, A. and Nicholson, R. B., Precipitation Hardening, Prog. in Mat. Sci., 10, 3 (1963).
11. Hornbogen, V. E., Aluminum, 43, p 41-47, p 115 - 170 (1967).

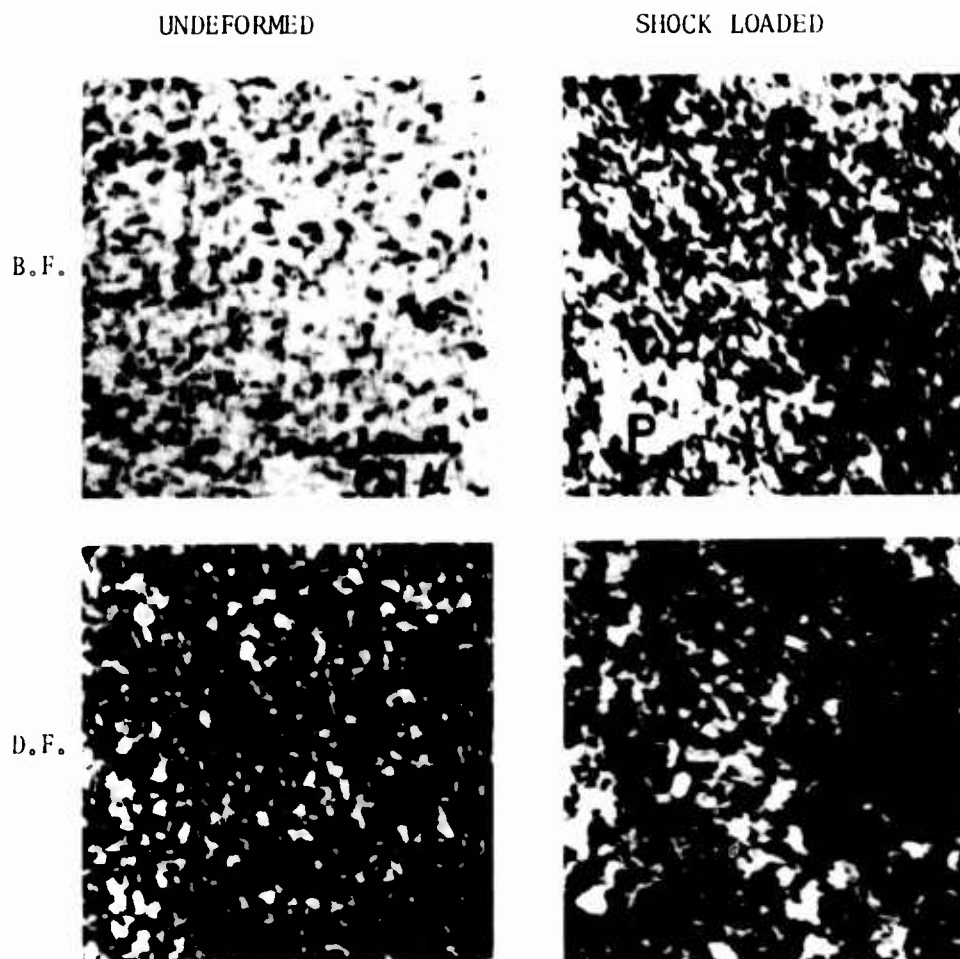


Figure 1. Transmission electron micrographs of samples from the under-aged Al-4 1/2 wt. % Cu alloy showing the same areas in bright field (B.F.) and dark field (D.F.), $\bar{g} = [200]$. All magnifications are identical. The shock loaded sample was taken from the area indicated by the arrow in Figure 3, approximately 2mm from the long axis.

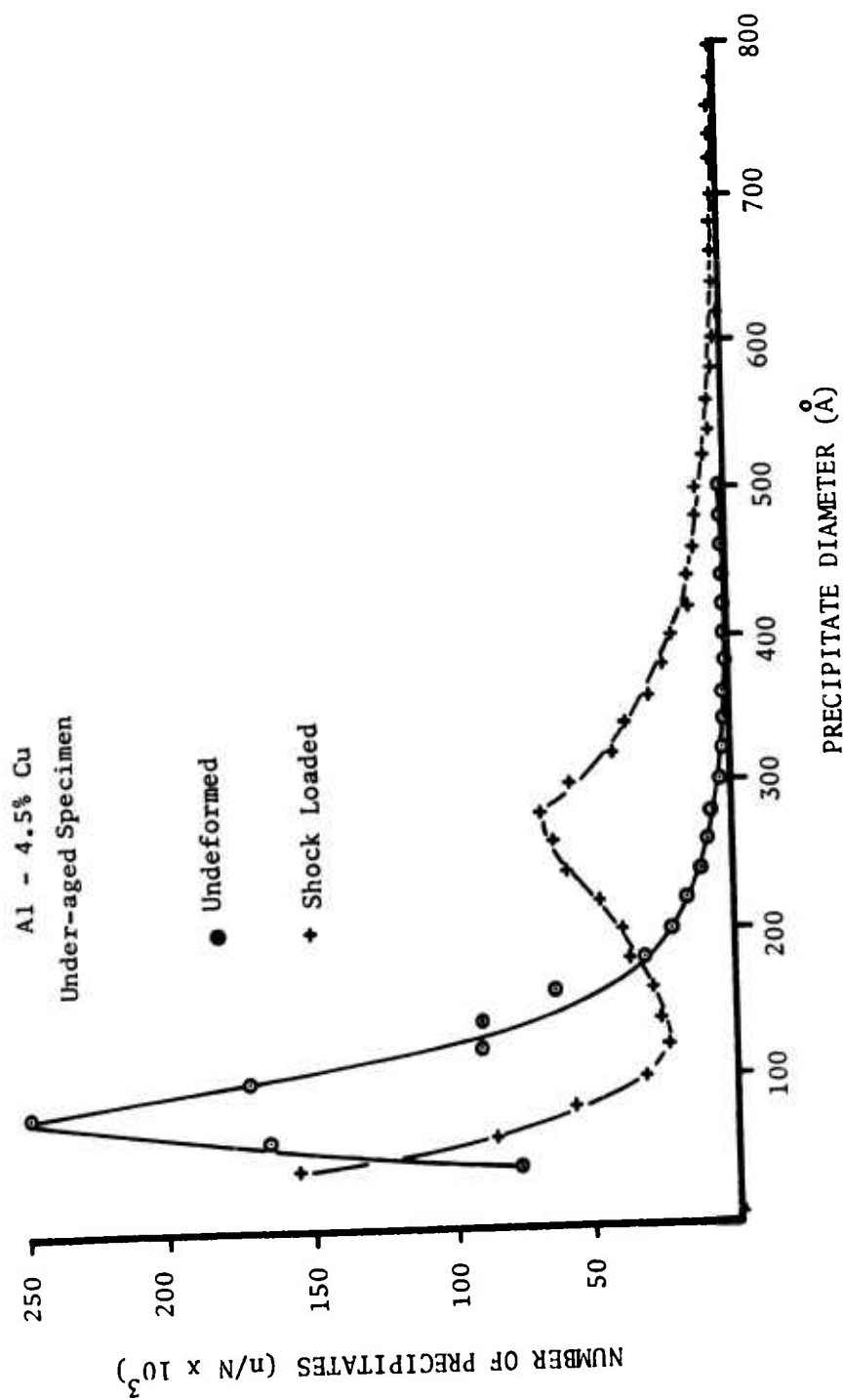


Figure 2. Normalized precipitate count indicating growth by ripening which occurs in the under-aged alloy during shock loading.

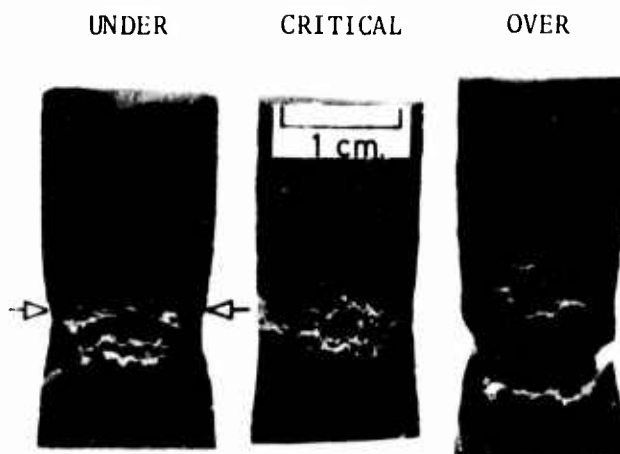


Figure 3. Macroscopic cross sections of under-, over-, and critical-aged samples after shock loading. The free surface of the specimen is the lower end shown in the figure.

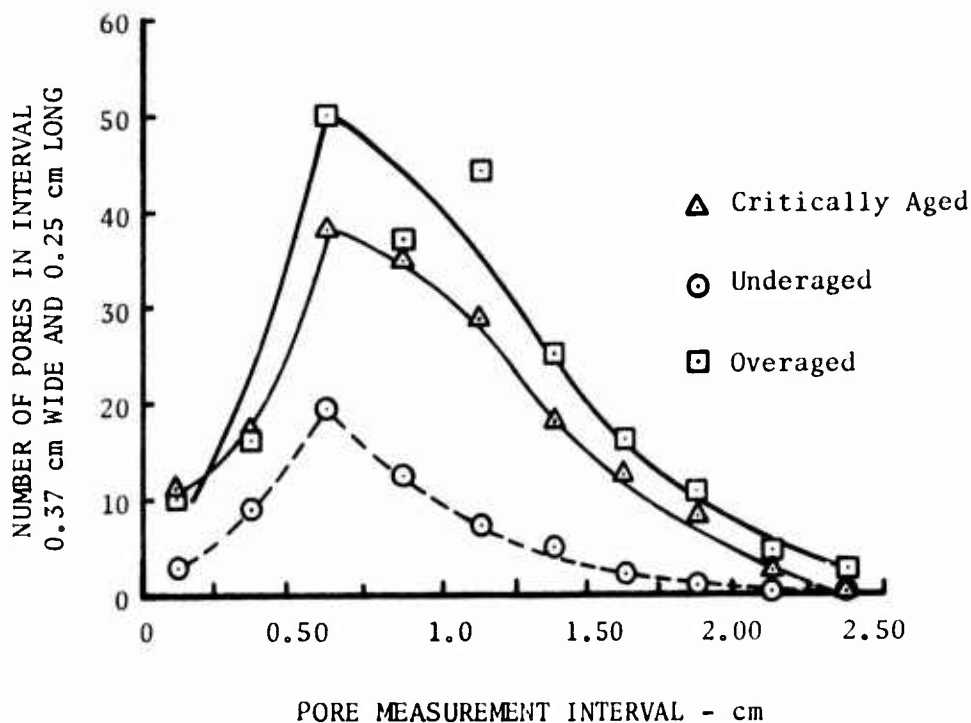


Figure 4. Number of pores in the 25 to 50 micron size range counted in areas 0.37 cm wide x 0.25 cm long versus counting area location, measured from the free surface.

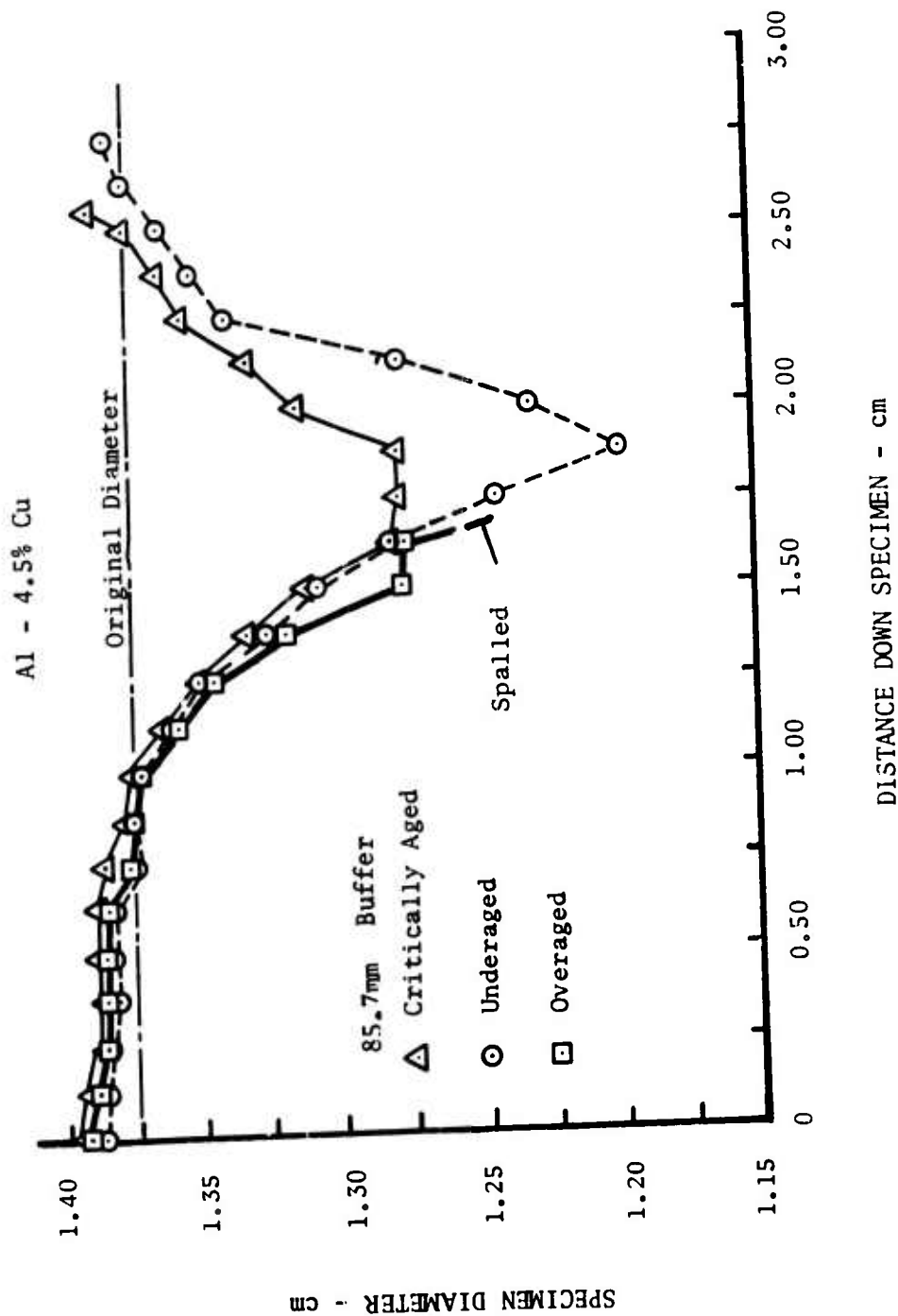


Figure 5. Necking response of shock loaded samples. Final diameter is plotted versus length from loaded end. Original diameter is shown by the dashed line.

SIMULATION OF A SIMPLE LORENTZ PLASMA WITH
A RANDOM DISTRIBUTION OF INDUCTIVELY-LOADED DIPOLES

JOHN F. W. DIETZ, MR.
*GEORGE MERKEL, Ph.D.
HARRY DIAMOND LABORATORIES
WASHINGTON, D. C.

1. INTRODUCTION

The complete simulation of the electromagnetic pulse, EMP, associated with a nuclear burst involves a number of interrelated parameters. The most obvious parameters are the waveform and the magnitude of the EMP. However, if the system being subjected to the nuclear EMP is in a preionized region such as the ionosphere, we would also have to simulate the ionization or dielectric constant of the medium surrounding the system such as a missile or satellite. In this work we investigate the feasibility of using an artificial dielectric consisting of a random distribution of inductively-loaded short dipoles to simulate the macroscopic electromagnetic properties of a simple Lorentz plasma.^{1,2}

A review of the literature reveals a number of papers on the plasma simulation properties of artificial dielectrics consisting of a rigid cubic lattice of three-dimensional grids.¹⁻³ Experimental work, reported in the literature, has been done only on two-dimensional grid lattice structures.¹⁻⁴ That a periodic grid structure would produce band structure resonances analogous to Bragg scattering has been discussed from a theoretical viewpoint, but has not been experimentally investigated.^{5,6}

The rigid cubic grid lattice structure has the disadvantage of being somewhat unwieldy and difficult to support and fit around an object with curved surfaces. In this paper we describe another approach to the simulation of a Lorentzian plasma. We propose a granular pellet-like artificial dielectric consisting of a random distribution of styrofoam spheres containing inductively-loaded dipoles which can yield macroscopic electromagnetic constitutive relationships

similar to those of a plasma.

A Lorentz plasma can be represented from the viewpoint of macroscopic electromagnetic theory, as a lossy dielectric with a complex dielectric constant given by^{1,2}

$$\epsilon_p = \epsilon_0 \left[1 - \frac{\omega_p^2}{(\nu^2 + \omega^2)} + j \frac{\omega_p^2 (\nu/\omega)}{(\nu^2 + \omega^2)} \right], \quad (1)$$

where ϵ_0 is the dielectric constant of free space, ω is the angular frequency of the electromagnetic field, ω_p is the plasma frequency, and ν is the collision frequency between electrons and the gas molecules. We propose to simulate this macroscopic dielectric constant with an artificial dielectric consisting of a random distribution of inductively-loaded dipoles encased in styrofoam pellets. A schematic drawing of such a dipole is shown in figure 1. The quantities 2ℓ , L , and R are the length, inductance, and resistance of the loaded dipole, respectively. As we will show, using simple quasi-static arguments, if the load impedance ωL is much greater than the capacitance impedance of the dipole, the permittivity of a random distribution of inductively-loaded dipoles is

$$\epsilon_p = \epsilon_0 \left\{ 1 - \frac{\frac{4}{3} \frac{\ell^2}{\epsilon_0} \frac{N}{L}}{\left(\frac{R}{L}\right)^2 + \omega^2} + j \frac{\frac{4}{3} \frac{\ell^2}{\epsilon_0} \frac{N}{L} \left(\frac{R}{L}\right) \frac{1}{\omega}}{\left(\frac{R}{L}\right)^2 + \omega^2} \right\}, \quad (2)$$

where N is the number density of the dipole pellets. One can see by comparing equation (1) with equation (2) that we can simulate a plasma with plasma frequency ω_p and collision frequency ν by setting

$$\frac{4}{3} \frac{\ell^2}{\epsilon_0} \frac{N}{L} = \omega_p^2 \quad (3)$$

and

$$\frac{R}{L} = \nu. \quad (4)$$

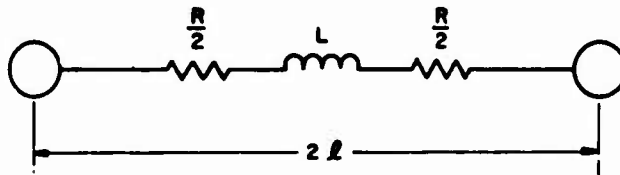


Figure 1. Inductively loaded dipole.

2. THEORETICAL DISCUSSION

2.1 Quasi-Static Approach

We will first approach the problem of our proposed artificial dielectric from a quasi-static viewpoint, assuming that the inductive load of the dipole is much larger than the driving point capacitive impedance of a short dipole. In a subsequent section we will explicitly consider the dipole capacitance.

We want to construct an artificial dielectric from a pellet-like medium that has an index of refraction given by¹

$$n = \left(\frac{\epsilon_p}{\epsilon_0} \right)^{1/2} \quad (5)$$

and an intrinsic impedance given by

$$\eta = \left(\frac{\mu_0}{\epsilon_p} \right)^{1/2}, \quad (6)$$

where ϵ_p is given by equation (1).

In general the index of refraction of an artificial dielectric may be calculated in a manner analogous to that employed in calculating the index of refraction of a molecular medium. Assuming that the random obstacles are not too closely packed so that we do not have to resort to the Clausius-Mossotti relation, the index of refraction is given by⁷

$$n = [\epsilon_1 (1 + N\chi_e/\epsilon_0) \mu_1 (1 + N\chi_m/\mu_0)]^{1/2} \quad (7)$$

where

- N = Number of scattering obstacles per unit volume,
- χ_e = Electric polarizability of a scattering obstacle,
- χ_m = Magnetic polarizability of a scattering obstacle,
- c_1 = 1 (for styrofoam),
- μ_1 = 1 (for styrofoam),
- ϵ_0 = permittivity of free space,
- μ_0 = permeability of free space.

A promising candidate for a scattering obstacle is an inductively-

loaded electric dipole. Consider the very idealized inductively-loaded dipole schematically shown in figure 1. Let an electric field E be applied parallel to the line between two spheres separated by the distance 2ℓ . The field E then establishes a voltage $V=2\ell E e^{j\omega t}$ between the spheres. Assume that the inductance between the two spheres is L , and also assume that $L\omega \gg 1/\omega C$, where C is the capacitance between the two spheres. Since we are dealing with a time harmonic field, assume that the charge accumulated on the dipole spheres is given by

$$q = Qe^{j\omega t} . \quad (8)$$

Then

$$2\ell E e^{j\omega t} = LQ(-\omega^2)e^{j\omega t} + RQj\omega e^{j\omega t} ,$$

so that

$$q = - 2\ell E e^{j\omega t} / (L\omega^2 - jR\omega) . \quad (9)$$

The dipole moment along the axis of the inductive dipole in figure 1 is then

$$p = 2\ell q = - 4\ell^2 E e^{j\omega t} / (L\omega^2 - jR\omega) . \quad (10)$$

If we wish to calculate the average polarization in the direction of the incident electric field when the field is not parallel to the dipole, we must multiply equation (10) by $\cos^2\theta$ and integrate over all possible values of θ .⁸

Then $p_{av} = \frac{1}{3} p$. Since $p_{av} = \chi_{eav} E e^{j\omega t}$ the polarizability, χ_{eav} , is given by

$$\chi_{eav} = - 4\ell^2 / 3(L\omega^2 - jR\omega) , \quad (11)$$

and, if $\chi_m \approx 0$,

$$\epsilon_p = \epsilon_0 [1 - 4\ell^2 N / 3\epsilon_0 L \omega^2 (1 - jR/\omega L)] \quad (12)$$

$$= \epsilon_0 \left\{ 1 - \frac{\frac{4}{3} \frac{\ell^2 N}{\epsilon_0 L}}{\left(\frac{R}{L}\right)^2 + \omega^2} + j \frac{\frac{4}{3} \frac{\ell^2 N}{\epsilon_0 L} \left(\frac{R}{L}\right) \frac{1}{\omega}}{\left(\frac{R}{L}\right)^2 + \omega^2} \right\} . \quad (13)$$

Finally, we note that if equations (3) and (4) are satisfied, we have a one-to-one correspondence between equations (1) and (13).

2.2 Explicit Consideration of Dipole Capacitance

Up to this point we have assumed that the inductance used to load the short dipole dominated the behavior of the dipole. Using the

results of King⁹ and of Harrington¹⁰, we can show that the electric dipole moment of a short inductively-loaded dipole that is aligned parallel to an incident electric field E is given by

$$p = \frac{-j B^2 E e^{j\omega t}}{4\omega(j\omega L + R + Z_{22})} \quad (14)$$

where B is the dipole length, $(j\omega L + R)$ is the impedance of the centrally located load and Z_{22} is the short dipole impedance. King gives the value of Z_{22}

$$Z_{22} = \frac{\eta}{2\pi} \left[\frac{\left(\frac{\omega}{c}\right)^2 B^2}{12} - j \frac{4 \ln\left(\frac{B}{A}\right) - 6.78}{\left(\frac{\omega}{c}\right) B} \right] \quad (15)$$

where η is the impedance of free space, $120\pi\Omega$, and A is the dipole radius. As before

$$p_{av} = 1/3 p,$$

so that

$$\chi_{av} = \frac{-j B^2}{12\omega(j\omega L + R + Z_{22})} \quad (16)$$

If we set $Z_{22} = 0$, $B = 2\ell$, and end load the dipole so that the current distribution is uniform and not triangular, we obtain the same result as that given by equation (11)¹¹.

The appearance of Z_{22} in the denominator of expression (14) is inconsistent with our desire to model a simple Lorentzian plasma. Specifically, when the frequency is such that the inductive reactance of the load equals the capacitive reactance of the dipole, the artificial dielectric will pass through a resonance. We can use the inductively-loaded dipoles to simulate a plasma only if $\omega L \gg Z_{22}$. We will discuss the effect of Z_{22} in a subsequent section. An insight into the limitations imposed by the dipole self impedance can be obtained by calculating χ_{av} for some realistic frequencies and dipole parameters.

2.3 Mutually Perpendicular Dipole Scatterers

Our arguments presented thus far have assumed that the distance between the inductively-loaded dipoles is large enough that their interaction can be neglected. The use of equation (7) is based on the assumption of negligible interaction between scattering objects. A way of increasing the value of N in (7) by a factor of 3 without violating the assumption of negligible interaction of neighboring

scatterers is to construct each scatterer out of three mutually-perpendicular, inductively-loaded dipoles. If we substitute three mutually-perpendicular dipoles for the single randomly-distributed dipoles, we simply multiply the average value of the polarizability given in equation (16) by 3.

By constructing our scatterers out of three mutually-perpendicular, inductively-loaded dipoles we accomplish more than simply increasing the number of scatterers by three: the polarizability of the scattering centers is made to be independent of scatterer orientation. To be specific, the scatterer, consisting of three identical, short, mutually-perpendicular dipoles centered at the origin with a dipole along each axis, can be described with the following simple dyadic:

$$\underline{\chi} = \chi (\hat{i} \hat{i} + \hat{j} \hat{j} + \hat{k} \hat{k}) = \chi \underline{I}$$

The dipole moment induced by an arbitrary field \vec{E} is then

$$\begin{aligned} \vec{p} &= \underline{\chi} \cdot \vec{E} \\ &= \underline{\chi} \cdot E (\hat{i} \sin \theta \cos \phi + \hat{j} \sin \theta \sin \phi + \hat{k} \cos \theta) \\ &= \chi E (\hat{i} \sin \theta \cos \phi + \hat{j} \sin \theta \sin \phi + \hat{k} \cos \theta), \end{aligned} \quad (17)$$

but $(\sin \theta \cos \phi)^2 + (\sin \theta \sin \phi)^2 + \cos^2 \theta = 1$. Therefore, the magnitude of the dipole moment is χE , and its direction is in the direction of the electric field regardless of the orientation of the scatterer.

2.4 Coupling of Scatterers

Thus far, we have considered our scattering dipoles to have a spatial distribution such that their mutual interactions can be neglected, i.e., we have assumed that the local field at a dipole scatterer is given by the applied field. As a first approximation to the effect of the other scatterers upon the local field of a scatterer, we can combine the expression for the polarizability, $\chi_{av3} = 3 \chi_{av}$, with the well-known Clausius-Mossotti equation.¹²

The permittivity of our artificial dielectric is then given by

$$\epsilon = \epsilon_0 \left[\frac{1 + 2 N \chi_{av3}/3\epsilon_0}{1 - N \chi_{av3}/3\epsilon_0} \right]. \quad (18)$$

When the number density of scatterers N is small, and $\chi_m \approx 0$, equations 7 and 18 are equivalent. Strictly speaking, equation (18) is based upon the assumption of a cubic lattice spatial distribution, and the applicability of (18) to an accumulation of dipole scatterers that have been dumped into a container must be justified.

2.5 Deviations from the Clausius-Mossotti Equation Due to Randomness of Scatterers

J. G. Kirkwood¹³ has calculated the deviations that one might expect from the Clausius-Mossotti equation in the case of a random distribution of molecules consisting of hard spheres. Specifically, Kirkwood considered a molecular interaction potential of the form

$$W(r) = \begin{cases} \infty & 0 \leq r \leq a \\ 0 & r > a \end{cases} \quad (19)$$

where a is the diameter of a molecule.

Kirkwood's model is not wholly applicable to an artificial dielectric formed by dumping spherical scatterers into a container because the scatterers do not necessarily assume a completely random distribution.

It is interesting to note that statistical mechanical averaging based on the molecular interaction equation (19) yields a value of the permittivity that is not a function of the thermodynamic temperature T . This follows because the value of the exponential term $\exp[-W(r)/kT]$ is either 0 or 1. In a gas or fluid the averaging can be over time; in our static collection of scattering dipoles the averaging would have to be over an ensemble of different containers, each filled with an aggregation of spherical dipole scatterers. Kirkwood's approach can be summarized as follows: As in the lattice case, the local field is obtained by summing over the potentials due to individual dipoles. We can say that the effective polarizing field E_{loc1} at scatterer 1 will be given by $E + E_1$, where E is the external applied field and E_1 is given by the sum over N individual dipoles:

$$\vec{E}_1 = - \sum_{\substack{k=1 \\ k \neq 1}}^N \vec{T}_{1k} \cdot \vec{p}_k \quad (20)$$

where the dipole-interaction dyadic is given by:

$$\vec{T}_{1k} = \frac{1}{4\pi\epsilon_0} \frac{1}{r_{1k}^3} \left[\underline{I} - 3 \frac{\vec{r}_{1k} \vec{r}_{1k}}{r_{1k}^2} \right] \quad (21)$$

The polarization of an individual scatterer is given by

$$\vec{p}_1 = \epsilon_0 \underline{\alpha} \cdot \vec{E}_{loc1} = \epsilon_0 \underline{\alpha} \cdot (\vec{E} + \vec{E}_1) .$$

We therefore have N simultaneous equations

$$\vec{p}_1 + \alpha \epsilon_0 \sum_{k=1}^N \vec{T}_{1k} \cdot \vec{p}_k = \alpha \epsilon_0 \vec{E} \quad i = 1, \dots, N$$

that would have to be solved in order to obtain E_{loci} for each of the individual scatterers in the distribution of N scatterers in volume V . We can average the N equations and obtain

$$\vec{p} + \alpha \epsilon_0 \sum_{k=1}^N \vec{T}_{1k} \cdot \vec{p}_k = \alpha \epsilon_0 \vec{E} \quad (22)$$

Kirkwood¹³ introduced the following fluctuation term

$$\vec{n} = \sum_{i=1}^N \sum_{i \neq k} (\vec{T}_{ik} \cdot \vec{p}_k - \vec{T}_{ik} \cdot \vec{p}_i) \quad (23)$$

and obtained

$$\vec{p} + \alpha \epsilon_0 \sum_{k=1}^N \vec{T}_{1k} \cdot \vec{p} + \vec{n} \alpha \epsilon_0 = \alpha \epsilon_0 \vec{E} \quad (24)$$

If we can assume that

$$\vec{T}_{12} \cdot \vec{p}_1 = \vec{T}_{12} \cdot \vec{p}_2, \quad (25)$$

the fluctuation \vec{n} is equal to zero, and the permittivity ϵ is given by

$$\epsilon = \epsilon_0 \left\{ \frac{1 + \frac{2}{3} \frac{N}{V} \frac{\chi_{av3}}{\epsilon_0}}{1 - \frac{1}{3} \frac{N}{V} \frac{\chi_{av3}}{\epsilon_0}} \right\}. \quad (26)$$

Equation (26) is to be compared with equation (18). After considerable mathematical manipulation and a number of approximations, Kirkwood managed to evaluate equation (23) for a substance consisting of a random distribution of hard, noninteracting spheres of diameter a :

$$n \approx \left(\frac{1}{9b} - \frac{5}{48v} \right) \frac{N^2 \alpha^2 D}{v} \quad (\alpha = \chi_{av3}/\epsilon_0), \quad (27)$$

where

$$b = \frac{2\pi N a^3}{3} = 4 N \left(\frac{\pi a^3}{6} \right) = 4 \text{ (volume of spherical scatterers).}$$

To obtain a rough estimate of the validity of using the Clausius-Mossotti equation for a random distribution of styrofoam ball dipole

scatterers, we can compare the value of n given by equation (27) with the value of the first two terms of the expansion of equation (24)¹³

$$1 - \alpha \sum \bar{T}_{ik} = 1 - \frac{1}{3} \frac{N\alpha}{v}$$

For the experimental situation discussed in Section 3:

$$\frac{n}{1 - \alpha \sum \bar{T}_{ik}} = .22 ,$$

and we might expect the Clausius-Mossotti equation to be a reasonable first approximation. Actually, as we have pointed out, Kirkwood's statistical model is not necessarily applicable to our aggregation of dipole scatterers because the dipole scatterers in the artificial dielectric do not necessarily assume a random distribution. The validity of the Clausius-Mossotti equation as applied to our artificial dielectric should really be tested empirically.

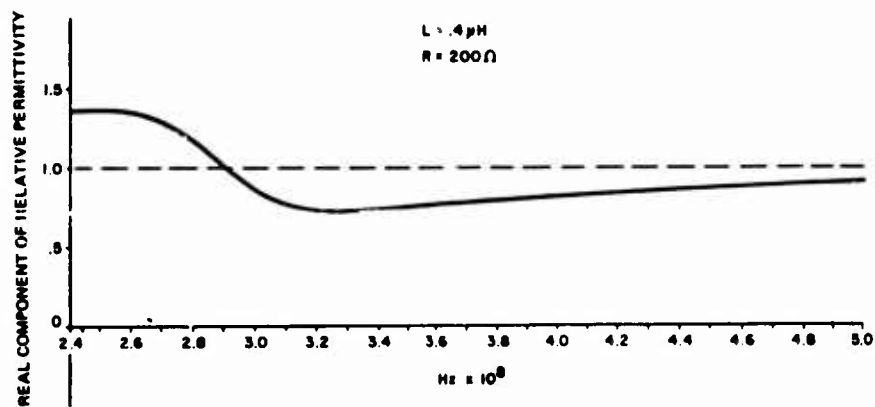
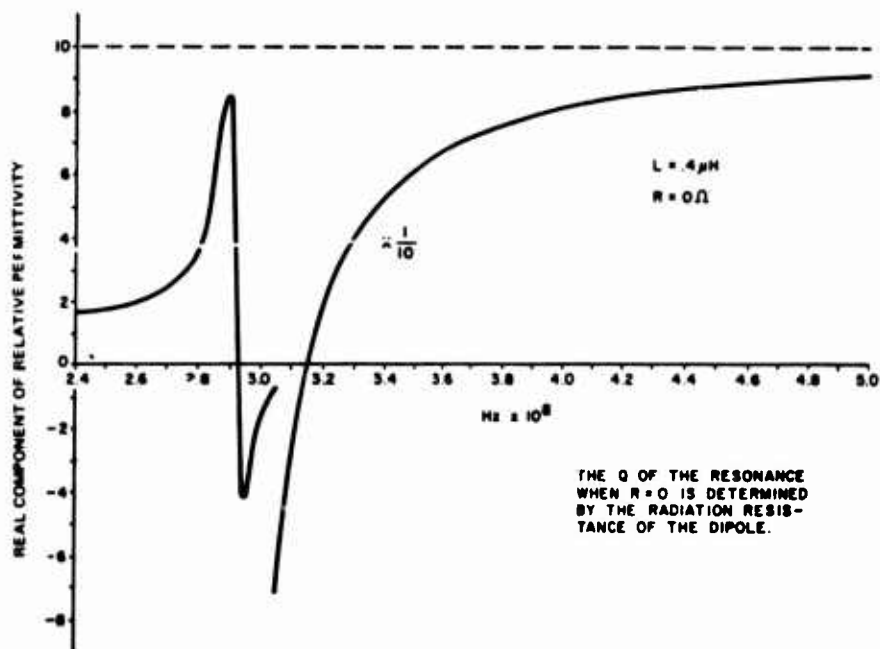
3. COMPARISON OF EXPERIMENT WITH THEORY

In order to orient ourselves with regard to the parameters such as the dipole load impedance and Z_{22} , we will use equation (18) and the expression for χ_{av} obtained in section 2.4 to design an artificial dielectric with some actual experiments in mind. The frequency range considered will be dictated by the band pass characteristics of the type 2300 waveguide. The type 2300 waveguide, the largest standard waveguide, has internal dimensions of 23.0 x 11.5 in., a lower cutoff frequency of 2.56×10^8 HZ, and a recommended upper frequency limit of 4.9×10^8 HZ. The frequencies in this range are much higher than the values usually quoted for nuclear EMP pulses, but we will consider designing an artificial dielectric to operate at these high frequencies because of the convenience of using waveguide techniques in measuring the properties of an artificial dielectric.

Let us consider the following parameters: we construct the artificial dielectric out of scatterers consisting of three mutually-perpendicular, 7-in. long dipoles loaded with an inductance of .4 μ H. The stems have a radius of .1 cm. We choose the density of scatterers to be 264 per cubic meter. This density corresponds to an aggregation of 48 scatterers throughout a rectangular container with dimensions 23 x 11.5 x 41.5 in.

In figures 2a and b we present values of the real part of the permittivity calculated with equation (18) with $L = 0.4 \mu$ H. Figure 2a, corresponding to a purely inductive load, emphasizes the resonance behavior of the artificial dielectric produced when the load inductance L is in resonance with the capacitive impedance of the dipole. As can be expected the Q of the resonance is reduced by including a resistance in the dipole load impedance.

DIETZ & *MERKEL



Figures 2a and 2b. The real component of relative permittivity is calculated with equation (18) and

$$\chi_{av3} = \frac{-j B^2}{4\omega(j\omega L + R + Z_{22})}$$

In our first experiment we searched for the resonance predicted by equation (16) and shown in figure 2a. The experimental setup consisted of a sample of 48 7-inch diameter styrofoam balls placed in a large type 2300 waveguide with a matched termination. A series of V.S.W.R. measurements yielded a large resonance in the reflection coefficient at approximately the resonance frequency shown in figure 2a.

In a second series of experiments we measured the permittivity of our artificial dielectric utilizing the well-known shorted waveguide technique described in great detail by von Hippel.¹⁴ Very briefly, the technique consists of measuring the null point of a standing wave in a shorted waveguide. A sample of length l is then inserted into the shorted waveguide and the shift in the standing wave null is noted. The index of refraction of the dielectric sample of length l can then be calculated in terms of l and the shift in the standing wave null point.

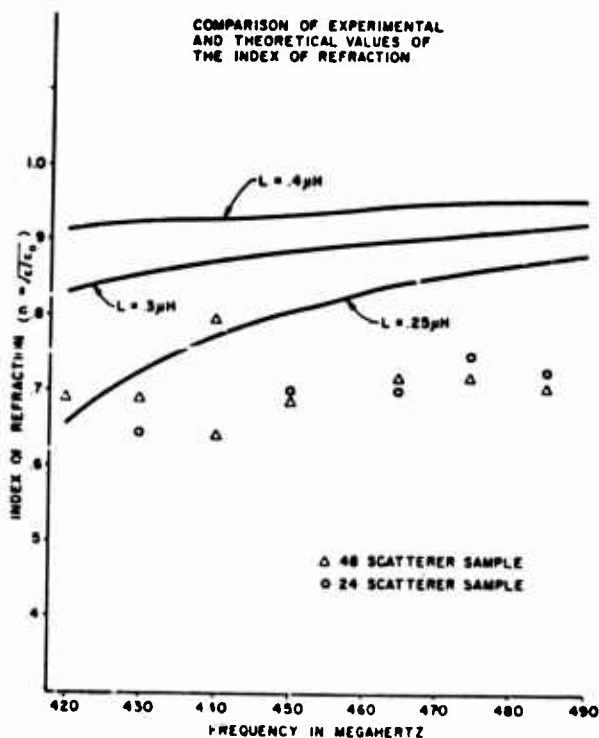


Figure 3. The accuracy of the experimental points is limited by the granularity of the sample and the corresponding indefiniteness of the sample edges.

Our experimental values of the index of refraction and some calculated values are shown in figure 3. The measurements were of two different artificial dielectric sample sizes; one sample consisted of an aggregation of 24 scatterers measuring 23 x 11.5 x 20.75 in. in volume. The length of the sample was $\ell = 20.75$ in. The other sample consisted of an aggregation of 48 scatterers in a rectangular volume measuring 13 x 11.5 x 41.5 in. In this case, $\ell = 41.5$ in.

The difference between the theoretical index of refraction curve corresponding to $L = 0.4 \mu\text{H}$ and $R = 0$ and the measured index of refraction is about 30 percent. A possible source of the disparity between experiment and theory is that we used the dimension of the cardboard box that enclosed the scatterers to obtain ℓ . The effective length of the very granular artificial dielectric sample is most probably not the size of the box containing the scatterers. The boundary of our artificial dielectric is somewhat nebulous and needs more study.

4. SUMMARY AND CONCLUSIONS

We first introduced a simple quasi-static model of the artificial dielectric consisting of inductively-loaded dipoles which (1) ignored the dipole capacitance, and (2) assumed that the dipole-dipole interaction could be neglected. These two simplifying assumptions yielded an expression for the permittivity of the artificial dielectric that behaved in a manner surprisingly analogous to the permittivity of a simple Lorentzian plasma. It was noted that the granularity of fluctuations of the artificial dielectric could be decreased by constructing the dipole scatterers out of three mutually-perpendicular scatterers.

Then the effect of the capacitance of the inductively-loaded dipole scatterers was examined and limits set on the assumption that the inductance dominated the behavior of the inductively-loaded dipole. Next, the effect of dipole-dipole interactions on the artificial dielectric was considered.

It is interesting that the more exact expressions for the dielectric constant of the artificial dielectric, which considered both the effect of the capacitive dipole impedance and the dipole-dipole interaction, did not model the behavior of a Lorentzian plasma as well as the original simplified approach. Nevertheless, the artificial dielectric did have an index of refraction less than one over a relatively broad spectrum.

The experimental examination of the theoretical expressions for the dielectric constant of our artificial dielectric was constrained by the dimensions and frequency characteristics of available waveguide equipment. Our predictions for the dielectric constant of the

artificial dielectric were found to be in fair agreement with experimental results.

5. ADDENDUM

Since the conclusion of the work described in the foregoing sections, we have checked some of the predictions of our closed form expressions, based upon our very simple model, with a relatively versatile and accurate computer code developed by Kuo and Strait¹⁵ which employs the method of moments as described by Harrington.¹⁶ The Kuo-Strait computer code calculates current distributions along arbitrarily-bent thin wires that result when the configuration of thin wires is subjected to an incident electromagnetic plane wave. The code also calculates scattering cross sections for arbitrary configurations. The application of Harrington's method of moments to the solution of the integro-differential equations that describe the interaction between our inductively-loaded, mutually-perpendicular scattering objects and an incident electric field can involve the computer inversion of rather large matrices when the inductor is treated not as a lumped load, but as a helix.

We used the Kuo-Strait code to examine the assumption that the current distribution along a dipole is triangular, which is an assumption that is employed to derive equations (14) and (15). We found that there is a slight deviation, approximately 10%, from a triangular current distribution at 400 megahertz when the inductively-loaded dipole is 7 in. long. The code was also used to check the conclusions reached by the arguments leading up to equation (17), i.e., that the polarizability of a scatterer consisting of 3 perpendicular dipoles is independent of scatterer orientation. We found that, for 7 in. long, perpendicular dipoles at 400 megahertz, there could be as much as a 15% change in back-scattering cross sections, depending upon the relative orientation of dipoles and the direction of the incident field. However, for 3.5 in.-long, perpendicular dipoles at 400 MHz, the back scattering cross section or polarizability of the scattering object was independent of dipole orientation if the dipoles were sufficiently close together to preserve symmetry. That is, for the shorter dipoles, the conclusions of equation (17) are correct.

The Kuo-Strait code has been used to design a follow-up experiment with shorter inductively-loaded dipoles. The object is to have a much larger number of scatterers or "molecules" per unit volume and, thus, have a much finer granularity. Reducing the size of the scatterers, however, creates new problems. The shorter dipole length produces an increase in the dipole impedance Z_{22} , so that the inductance must be increased to meet the requirement that $\omega L \gg Z_{22}$. Intuitively, one can increase the inductance of the lumped parameter choke by increasing the number of turns in the choke; but, if the choke length is held

fixed, the resulting decrease in the coil separation will create an increase in the distributive capacitance of the inductor.¹⁷ The Kuo-Strait code was used to design the lumped inductance so that the inductance was not swamped by distributive capacitance. In our final design the total dipole length was 5 cm, and the inductor length was 1 cm. This shorter dipole will allow an increase in dipole number density by a factor of 30. Thus the number of scatterers in the volume described in section 4 will be increased from 48 scatterers to approximately 1500 scatterers.

As indicated in Section 3, it is difficult to define the boundary of the artificial dielectric. In our analysis of the experimental measurements, we have simply assumed that the edge of the artificial dielectric can be defined by an abrupt step function rather than a more realistic gradual transition. With a reduction in the size of our scatterers or molecules the assumption of a step function transition becomes more realistic. A possible phenomenological approach to treating the boundary of the artificial dielectric might be found in the nuclear optical model. In this model a nucleus (consisting of from approximately 20 to perhaps 230 nucleons) is interpreted as a "cloudy crystal ball with fuzzy edges". In this approach the edge of an artificial dielectric with permittivity $\epsilon(z)$ would be described with a gradual-variation Fermi gas-like function of the form¹⁸

$$\epsilon(z) = \frac{\epsilon_p}{1 + \exp[-(z-z_0)/\Delta]}$$

(Note that when Δ approaches 0, $\epsilon(z)$ approaches a step function at z_0 .) The fuzziness parameter, Δ , could be determined empirically to yield improved agreement between our theoretical predictions and our measured results.

6. REFERENCES

1. W. Rotman, "Plasma Simulation by Artificial Dielectric and Parallel-plate Media," IRE Transactions on Antennas and Propagation, Vol. AP-13, p. 587, July 1965.
2. K.E. Golden, "Simulation of a Thin Plasma Sheath by a Plane of Wires," IRE Trans. Ant. Prop., Vol. AP-13, p. 587, July 1965.
3. J. Brown, "Microwave Lenses," Methuen & Co., Ltd, London, 1953.
4. M.M.Z. El-Kharadly, "Some Experiments on Artificial Dielectrics at Centimeter Wavelengths," Proc. IEE, Paper No. 1700, Vol 102B, p. 22, January 1955.
5. L. Brillouin, Wave Propagation in Periodic Structures, McGraw-

Hill, New York, 1946.

6. H.S. Bennett, "The Electromagnetic Transmission Characteristics of Two Dimensional Lattice Medium," J. Appl. Phys., Vol. 24, pp. 785-810, June 1953.
7. H. Jasik, Antenna Engineering, McGraw-Hill Book Co., New York, 1963, Sec 14.6. Schelkunoff & Friis, Antennas Theory and Practice, John Wiley & Sons, Inc., New York, 1952, p. 577.
8. G. Merkel, "Simulation of a Simple Lorentz Plasma with a Random Distribution of Inductively-Loaded Dipoles", HDL-TR-1637, Harry Diamond Laboratories, Washington, DC, Nov. 1973.
9. R.W.P. King, The Theory of Linear Antennas, Harvard University Press, Cambridge, Mass., 1956, p. 184.
10. R. F. Harrington, "Small Resonant Scatterers and Their Use for Field Measurements," IRE Transactions of Microwave Theory & Techniques, May 1962, pp. 165-174.
11. R.W.P. King, op. cit. p. 470.
12. C. Kittel, Solid State Physics, John Wiley & Sons, New York, 1956, p. 163.
13. J.G. Kirkwood, "On the Theory of Dielectric Polarization," J. Chem. Phys., Vol. 6, p. 592, September 1936.
14. A. R. von Hippel, Editor, Dielectric Materials and Applications, The M.I.T. Press, Cambridge, Mass., 1954, Sec 2, p. 63, W. Westphal.
15. Dah-Cheng Kuo and Bradley J. Strait, "Improved Programs for Analysis of Radiation and Scattering by Configurations of Arbitrarily Bent Thin Wires," AFCRL-72-0051, Scientific Report No. 15, Jan 1972.
16. Roger F. Harrington, "Field Computation by Moment Methods," The McMillan Co., New York, 1968.
17. Frederick E. Terman, "Electronic and Radio Engineering", McGraw-Hill, New York, 1955, p. 31.
18. M. A. Preston, "Physics of the Nucleus", Addison-Wesley Co., Inc., Reading, Mass., 1962, pp. 544-545.

PHOTO- AND ELECTRIC FIELD-EFFECTS IN ENERGETIC MATERIALS

Dr. David S. Downs, Dr. Wayne Garrett, Dr. Donald A. Wiegand,
Dr. Thaddeus Gora, Mr. Marcel Blais, Mr. Arthur C. Forsyth,
and Dr. Harry D. Fair Jr.
Explosives Division, Feltman Research Laboratory
Picatinny Arsenal, Dover, N.J.

INTRODUCTION

At the 1972 Army Science Conference, Fair et al. (1) indicated the possibility that electronic processes in explosive solids could be used to achieve the direct, non-thermal initiation of both primary and secondary explosives. The theoretical basis for the relationship between chemical instability and electronic structure was reviewed, and it was experimentally demonstrated that a number of azide compounds could be initiated by purely electronic means. A combination of a low intensity static electric field and optical excitation in particular spectral regions was found to result in initiation (the photoelectronic initiation or PEI effect). Neither stimulus alone (at the intensities used) produces initiation, and it was shown that the effect could not be explained by purely thermal processes.

A detailed understanding of the microscopic processes involved in the PEI effect in azide compounds would be an important step toward the ultimate goal of achieving the electronic initiation of secondary explosives. Yet this is a monumental undertaking, for two important complications must be overcome: The first involves the nature of the experiments necessary. A large number of parameters must be examined, and their effects characterized; but each data point consumes at least one single crystal sample. The growth and preparation of these samples is difficult, time-consuming and hazardous. Secondly, the phenomenon being examined is very complex. It involves the simultaneous, time-dependent interaction of effects which can be characterized as electronic (electronic excitation, photoconductivity, perhaps carrier injection and multiplication), chemical (various decomposition processes), and material transport (leading to gas evolution and metal

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

colloid formation), and lead ultimately to a thermal and/or mechanical takeover involving self-sustaining exothermic reactions and detonation.

We report here our recent theoretical and experimental results which lead toward an understanding of the mechanisms involved in the PEI effect in lead azide single crystal samples. Our experimental studies have largely focussed on isolating the effects of optical excitation from those of electric field. Nitrogen gas and metallic lead are the ultimate products of lead azide decomposition. Hence decomposition in the spectral region appropriate to the PEI effect was monitored by directly studying the resulting gas evolution; and changes in the optical absorption were studied as a means of detecting colloidal lead formation. Also reported are electric field and electrode effects on initiation (without simultaneous optical excitation). From the results of these investigations we propose a model to account for the PEI effect.

BACKGROUND

A. Theoretical

A general theory relating electronic structure to chemical instability was developed by Williams (2). Briefly, the quantum-mechanical Hamiltonian governing the motion of the nuclei (atoms and/or molecules) in a solid contains an effective potential in terms of internuclear distance \underline{R} .

$$V_{\text{eff}}(\underline{R}) = V(\underline{R}) + E_{\text{el}}(\underline{R});$$

where $V(\underline{R})$ is the interaction potential between nuclei, and $E_{\text{el}}(\underline{R})$ contains the dependence of electronic energies on internuclear separation. As the latter term can vary strongly from one electronic state to another, the internuclear forces governing chemical reactions and therefore decomposition processes can be strongly affected by populating excited electronic states, as shown in Fig. 1.

An explosive solid differs from a reactive gaseous mixture in several important respects. The constituent atoms or molecules are densely packed, leading to their rather low mobility. Excited electronic states in a solid may differ substantially from the excited states of its constituent atoms or molecules, and in some cases are spatially delocalized over hundreds of atomic or molecular volumes. These states can be excited by optical irradiation of appropriate energy, or by charge injection from appropriate contacts. They may be involved in long-term energy storage. And, perhaps most importantly, the

DOWNS, GARRETT, WIEGAND, JORA,
BLAIS, FORSYTH and FAIR

excited electronic states may be mobile in the solid; if they involve charge carriers, external electric fields can influence energy transport, and perhaps lead to avalanching effects.

B. Electronic Structure of Lead Azide

The electronic structure and thermochemistry of azide compounds, and of lead azide in particular, have recently been elucidated (1,3). Lead azide can be viewed as consisting essentially of Pb^{++} and N_3^- ions. The ultimate source of its reactivity is the highly exothermic nature of the conversion of azide ions to nitrogen gas. Thus at some stage, charge transfer is important in the PEI effect as well as in other initiation modes.

Lead azide has an energy band gap of approximately 4eV (3). The highest filled valence band most likely consists of electronic states of N_3^- , and the (unoccupied) conduction band of Pb^{++} states. A band gap transition is described as exciting an electron from an N_3^- state to a Pb^{++} state. Lower energy states of excitation are known to exist experimentally (1,3), and have been attributed to either of several types of excitons. While their nature has not been definitively established, the evidence to date appears to favor cationic excitons, consisting of excited states of lead ion. These general features of the electronic structure are, of course, responsible for the manner in which lead azide absorbs light. Figure 2 shows the absorption spectrum of a lead azide thin film sample, as well as the absorption edge for a single crystal sample (3,4).

C. Previous Photodecomposition and Field Initiation Results

Lead azide is known to decompose on exposure to ultraviolet light, giving off nitrogen gas and leaving ultimately a residue of metallic lead or mixtures of lead salts (such as lead oxides and carbonates). With moderate irradiation intensities the decomposition proceeds slowly. With extremely high intensities lead azide detonates and the reaction is completed in a matter of nanoseconds.

All previous photodecomposition studies have been performed on powder and thin film samples, using mainly 253.7nm irradiation. However, the absorption spectrum of lead azide commences at much longer wavelengths in the region of 400nm (refer to Fig. 2). Further, the use of powder and thin film samples has made the data difficult to interpret, and often not reproducible because of the contributions of surface, interface, particle size and defects associated with powders and films. Hence to understand these phenomena it was necessary to examine the spectral response and intensity

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

dependence of photodecomposition efficiency for single crystal samples of lead azide.

No previous attempts have been made to examine what happens to lead ions during photodecomposition. For reasons of overall charge neutrality, either neutral lead or lead salts are expected to form. A study of changes in the absorption spectrum as a function of prior irradiation and history was hence undertaken in an attempt to observe colloidal lead particles, again using single crystal samples.

Finally, a number of field initiation studies have been reported (5) for lead azide. All such studies to date have been performed on pressed pellet and powder samples and the results have therefore varied with particle size distribution and overall density. A number have involved at least one electrode not in contact with the sample; thus air breakdown and corona discharges were important in their interpretation. Studies to date have yielded little information on electric field effects in the bulk of a lead azide sample. To gain an understanding of these complex phenomena it was necessary to determine the condition under which electric fields induce initiation of single crystals of lead azide for a particular electrode material and sample-electrode geometry. Our configuration and contact choices were made to maximize the chances for current injection and minimize surface effects.

EXPERIMENTAL TECHNIQUES & RESULTS

A. Crystal Growing

The most critical obstacle to the experimental study and understanding of electronic initiation has been the lack of single crystals of sufficient size, purity and optical quality. Previous efforts to grow lead azide crystals gave rather discouraging results, with detonations of the crystal-growing solutions resulting in the loss of crystals and of equipment as well. Thus the task of developing a safe reliable method of crystal growth was undertaken using recent developments in the solution chemistry of azides, along with modifications of the techniques of crystal growth from solutions used for other compounds.

The lead azide crystal growing apparatus (6) is shown in Fig. 3. To limit the amount of explosive material in any one container and to ensure against the loss of hydrazoic acid at high temperatures, pyrex test tubes with screw caps containing silicone rubber gaskets were used as the growing vessels. To hold the test tubes inside the polycarbonate holder also served to

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

isolate the test tubes from each other should explosions occur. A proportional temperature controller with a platinum sensor and a 200 watt immersion heater maintained temperature control to within 0.01°C . The crystals were grown by slowly cooling saturated ammonium acetate solutions from 70°C at a rate of $0.5^{\circ}\text{C}/\text{day}$.

Harvesting the crystals is a critical step. Spontaneous explosions occur if all the crystal-growing solution is not immediately washed from the crystal surfaces. The best procedure involves pouring the solution and crystals from the test tube onto a porous cloth covering a ceric ammonium nitrate solution and immediately rinsing with warm distilled water. The azide remaining in solution is thus destroyed in the ceric solution.

X-ray precession photographs show that the resulting optically clear crystals are orthorhombic, the alpha-form of lead azide. The diffraction spots are sharp and there is no evidence of twinning. The largest crystals have approximate dimensions of $1.0 \times 0.5 \times 0.5 \text{ cm}$. No attempt was made to grow larger samples because of the danger involved in handling them. However, for most of the optical and electronic measurements, techniques had to be developed to cut and polish the crystals. The as-grown crystals normally develop one or more large faces that can be conveniently glued to a holding rod (with beeswax). Plates can then be cut with reference to the natural face. The cutting is accomplished with a string saw using a dilute ceric ammonium nitrate solution as the chemical cutting agent. The resulting crystal plates are then polished on a slow speed polisher using a micro-polishing cloth and a polishing compound of medium to very fine grain size. Figure 4 shows typical plates cut from large crystals as well as two uncut crystals in the bottom row.

B. Photo-Effects

To understand the detailed nature of the effects of irradiation on lead azide, optical transmission and mass spectrometric techniques were utilized. The optical transmission measurements were performed on cut and polished lead azide single crystal plates mounted in a vacuum cryostat (pressures of 10^{-3} to 10^{-5} Torr) in a Cary 14 R spectrophotometer. The optical density or fraction of light absorbed in passing through the sample was measured as a function of wavelength or energy and was determined at sample temperatures of 12, 78 and 300K . Changes in the optical density were measured following irradiation of the lead azide sample with a high pressure mercury lamp (250 watts) and appropriate filters.

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

The evolution of gas from single crystals was measured in a Varian ultra-high vacuum system utilizing a Vac-Ion pump to achieve base pressures of 10^{-8} Torr. The samples were irradiated with monochromatic light from a 1000 watt Mercury-Xenon lamp through a quartz prism monochromator. The gas detected either with a residual gas analyzer mass spectrometer or a Bayard-Alpert type vacuum gage. With this technique a monolayer of decomposition from the surface of a single crystal could be detected.

The optical transmission studies (7) indicate that the change in optical density produced in lead azide exposed to blue and ultraviolet irradiation results in the formation of colloids of metallic lead. In Fig. 5, the experimentally observed increase in optical density (due to irradiation of a lead azide single crystal with ultraviolet light) is given as a function of wavelength. The calculated optical density of a lead azide sample containing spherical (up to about 10 nm dia) lead colloids, using the Mie Theory, is also shown in Fig. 5. The good agreement between the observed and calculated optical spectra, plus the lack of temperature dependence of the optical density in irradiated crystals, indicate that ultraviolet and blue light irradiation of lead azide single crystals produce lead colloids.

The efficiency (number of N_2 molecules evolved/incident photon/sec) of the evolution of nitrogen gas from a large single crystal plate of lead azide has been studied as a function of wavelength of the exciting light, and the results, together with the optical absorption of lead azide, are shown in Fig. 6. The kinetics and energetics of the reactions and the specific electronic processes and transitions giving rise to the decomposition have been discussed elsewhere (8). The major result for this work concerns the magnitude of the quantum efficiency for the evolution of nitrogen gas (less than 10^{-5}) near the onset of optical absorption. The spectral dependence of the efficiency closely follows the optical absorption.

C. Electric Field Effect and Initiation Studies

To gain an understanding of the effect of the electric field in the absence of photo-effects a detailed study of the current-voltage characteristics and thickness dependence was made. Gold electrodes were applied to both sides of single crystal lead azide plates by vacuum evaporation. These gold electrodes were then contacted by gold wires using dots of silver paint. Sample thickness varied from .019 to .076 cm in this series of experiments. The electrode diameter, 2.36mm, was the same on all samples. By keeping the evaporated electrode diameter large compared to the sample thickness, the results can be analyzed on the basis of one-dimensional

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

planar current flow (9). An advantage of the sandwich-type electrode configuration is that surface currents can then be essentially eliminated. The samples were potted in RTV, a polymeric electronic potting composition, which aids in prohibiting current flow by any path other than through the bulk of the sample.

The sample chamber was evacuated by a forepump with an in-line liquid nitrogen trap. With samples of the size used in these experiments, the chamber is not damaged by the detonation. Performing the experiments under this type of vacuum is sufficient to eliminate the shock wave resulting from sample detonation and any dependence of the results on atmospheric humidity conditions.

An inherent procedural difficulty arose because of the possibility that lead azide crystals subjected to high electric fields may experience localized decomposition or other damage prior to detonation. This could act as a memory of previous treatment and might influence both the conductivity and critical voltage for detonation. One solution would be to use a fresh crystal for each value of applied voltage. However, this procedure would require an unrealistic number of samples with accurately controlled thickness. To minimize memory effects, each sample was subjected to the same voltage history. Our procedure consists of one-minute-on, one-minute-off application of voltage, increasing the voltage in 100V increments. The procedure was continued until the sample detonated, with current being monitored continuously. A typical current-response is shown in Fig. 7.

The current has a negative transient when the voltage is turned off which is not shown. Seven samples of varying thickness were investigated in this manner. The voltage at which detonation occurred, (V_{DET}), is shown in Table I along with the current just prior to initiation, I . It should be noted that the detonation does not necessarily occur immediately upon application of the voltage. The delay time was observed to vary. Typical current-voltage characteristics are shown in Figs. 8 and 9 and reflect the non-ohmic nature of the Au/lead azide system.

In a very simplified picture the voltage applied to the crystal can under some circumstances produce a constant average electric field in the sample given by V/L , where L is the sample thickness. It is assumed here that the contacts to the crystal do not affect the field distribution in any way, that is, the contacts are ohmic. On the other hand, it is possible that the contacts to the sample are non-ohmic and restrict the flow of current. This kind of contact is called a blocking or rectifying contact; and, because it

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

can present more resistance to current flow than the bulk resistance of the sample, the field distribution may be altered considerably. Gold forms a non-ohmic contact with lead azide (10). The effectively higher resistance of the Au/lead azide interface can result in a significant voltage drop across this region, with considerably less voltage actually applied across the bulk of the sample.

The average fields for detonation, calculated for each crystal from V_{DET} divided by the sample thickness L , range from $3.02 \times 10^4 - 4.21 \times 10^4$ V/cm. Thus there is a threshold in this field range at which detonation occurs. The scatter in values could be accounted for by differences in the crystal-electrode interfaces or differences due to crystal orientation. The threshold field values for the four thinnest samples (.019 - .025 cm) are all higher than those for the thicker samples (.043 - .076 cm). This would be expected if effects resulting from exposure to electric fields tend to lower the field threshold since the thicker samples experienced longer cumulative exposure times to electric fields.

MODEL FOR THE PHOTOELECTRONIC INITIATION EFFECT

The photolytic and electric field studies have led to a number of specific and general conclusions on the individual phenomena investigated. Our purpose, however, was not only to understand the separate phenomena but to relate them to understand the nature of the more complex PEI effect. It is now possible to infer a mechanism for the PEI effect.

We first review the evidence pertinent to the mechanism. Neither the field alone nor the light alone, at the intensities used, can produce initiation by simple thermal effects (1). Similarly, the combination of light and electric field intensities used in the PEI effect are not sufficient to lead to initiation thermally (1,11). Even if all of the energy released by exothermic photodecomposition reactions is taken into account, the magnitude of the efficiency for photodecomposition is far too small to provide sufficient thermal energy to initiate fast decomposition. The exciting light does lead to the production of nitrogen gas and to the formation of lead colloids. The resulting decomposition may produce energy states which enhance reaction. In addition, the exciting light produces large changes in the electrical conductivity (up to four orders of magnitude) by the production of photocarriers (1). Initiation clearly occurs in response to an applied electric field, and an average field value has been determined experimentally.

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

In order to initiate lead azide by light alone, an intensity six orders of magnitude greater than that used in our experiments is required (1). However, with the low light intensity used in the PEI effect, the value of the applied voltage is of the order of half that required to initiate with electric field alone. Therefore the primary effect of the exciting light combined with the applied electric field is to change the electric field distribution in the lead azide. This accounts for the factor of two or more decrease in average electric field necessary to achieve initiation with light. Figure 10 depicts a uniform voltage drop, which by Maxwell's equations predicts a constant electric field and overall zero net charge. This approximates conditions in the bulk of the crystal in the absence of light, and is generally correct independent of the nature of the contact/sample interface. When the light is turned on, charge carriers of both signs are produced, either directly or through the intermediate step of exciton formation. The original applied electric field separates charge carriers of opposite sign. The resulting charge distribution alters the electric field distribution. The nature and degree of the charge separation are related to the carriers' conductivity properties, the intrinsic defect structure of the crystal, and possibly to the photo-produced defects. Colloidal particles can act as sources or traps for charge carriers, and will be present near the illuminated surface. In addition, the charge distribution can be strongly affected by the presence of blocking contacts which prohibit the flow of charge into and out of the sample. Charge can accumulate at the interface, producing high fields across the contact region, and considerably altering the field in the bulk. For example, the arbitrarily chosen charge distribution shown in Fig. 11 leads to the indicated voltage and field distributions. The important point is that upon illumination, the resulting photo-induced charge distribution leads to an increase in the electric field at some point in the sample by a factor of two or more above that which it had in the nonilluminated case. This increase in internal electric field is sufficient to produce initiation.

Although there is not yet a clear understanding of why a threshold electric field at a given region of the sample causes initiation, one plausible explanation is that the electric field produces an avalanching multiplication of charge carriers. This results in a region of dense electronic excitation leading to an ultimate exothermic process. This is further supported by the fact that low-intensity light alone does not produce these regions of dense electronic excitation.

DOWNES, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

CONCLUSIONS

Our theoretical studies (1) indicated the possibility of initiating decomposition and detonation by several radically new mechanisms. The photoelectronic initiation effect in lead azide was the first experimental observation of the initiation of detonation by purely non-thermal electronic processes. To understand this phenomenon, techniques were developed to grow, cut and polish large pure single crystals of lead azide. The separation effects of light and electric fields on lead azide single crystals were studied to determine the detailed mechanisms of the interactions. These were related to develop an understanding of the PEI effect in terms of threshold average electric fields. Illumination creates photo-carriers which change the electronic charge distribution within the lead azide, raising the electric field to the threshold for initiation. These results provide the understanding to apply the photoelectronic initiation effect to other technologically important explosives, and provide the rational basis for pursuing related new mechanisms for initiation.

ACKNOWLEDGEMENTS

We are grateful for helpful discussions with Profs. Peter Mark and F.E. Williams.

REFERENCES

1. H.D. Fair, D.S. Downs, A.C. Forsyth, M. Blais and T.F. Gora, in Proceedings of Army Science Conference, West Point (1972).
2. F.E. Williams in Chemical Dynamics, J.O. Hirshfelder and D. Henderson, Editors (Wiley-Interscience, New York 1971), p. 289.
3. D.S. Downs and T.F. Gora in High Energy Chemistry: Physics and Chemistry of the Inorganic Azides, H.D. Fair, Editor (Plenum, New York, in publication).
4. H.D. Fair and A.C. Forsyth, J. Phys. Chem. Solids 30, 2559 (1969).
5. See for example, H.S. Leopold, Naval Ordnance Laboratory Tech. Rept. 73-125 (1973) and M.S. Kirshenbaum, PATR 4559 (1973).
6. W.L. Garrett, Mat. Res. Bull. 7, 949 (1972).
7. D.A. Wiegand, PATR 4080 (1970).
8. W. Garrett in High Energy Chemistry: Physics and Chemistry of the Inorganic Azides, H.D. Fair, Editor (Plenum, New York, in publication).
9. M. Lampert and P. Mark, Current Injection in Solids (Academic Press, New York 1970).
10. D.S. Downs, PATR (in publication).
11. D.S. Downs, T.F. Gora and M. Blais, PATR (in publication).

DOWNS, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

TABLE I. ELECTRIC FIELD EFFECTS ON LEAD AZIDE

Sample	Thickness L(cm)	Current at Detonation (amp)	Voltage at Detonation V_{DET}	Threshold Field V_{DET}/L (Volts/cm)
3-72-30	.019	1.2×10^{-9}	800	4.20×10^4
3-72-31	.024	6.9×10^{-8}	1000	4.17×10^4
3-72-32	.022	7.0×10^{-9}	800	3.60×10^4
3-72-33	.043	1.6×10^{-8}	1300	3.02×10^4
3-72-34	.025	1.4×10^{-8}	1000	4.00×10^4
3-72-35	.076	5.5×10^{-9}	2700	3.54×10^4
3-72-36	.073	1.1×10^{-8}	2400	3.26×10^4

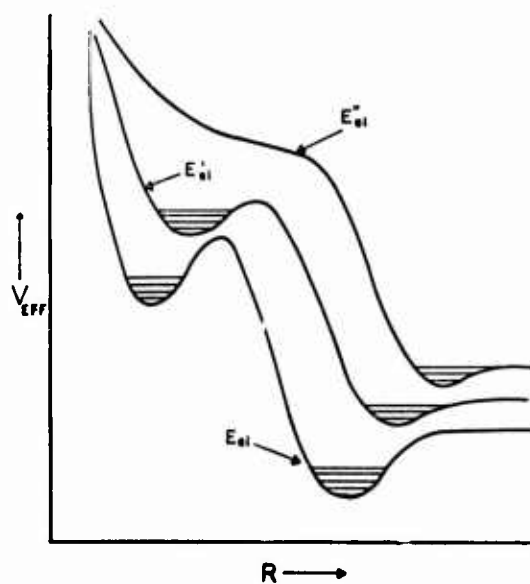


Figure 1. Reaction coordinate diagram depicting possible changes in activation energy due to excited electronic states.

DOWNES, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

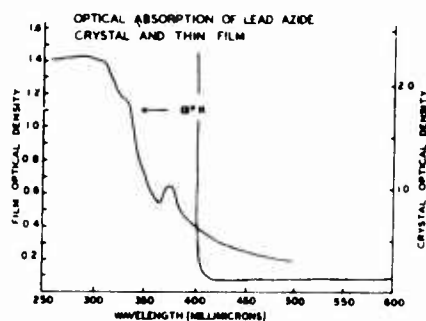


Figure 2. Comparison of optical absorption spectra of thin film and crystal.

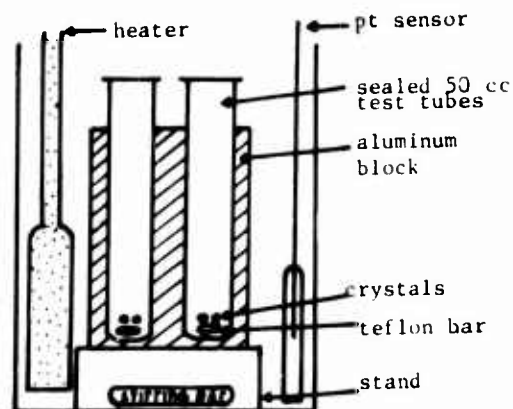


Figure 3. Crystal growing apparatus.

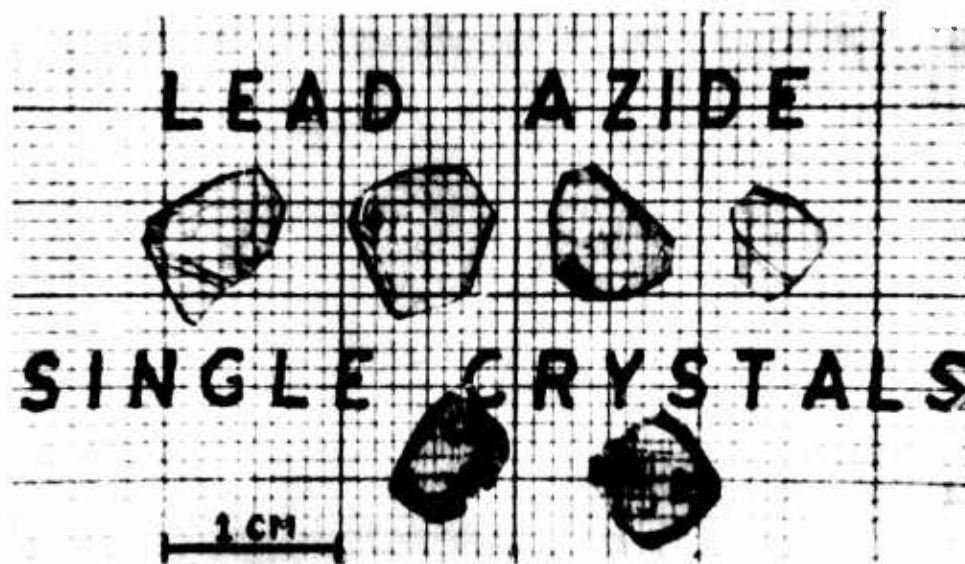


Figure 4. Solution grown lead azide crystals.

DOWNES, GARRETT, WIEGAND, GORA,
BLAIS, FORSYTH and FAIR

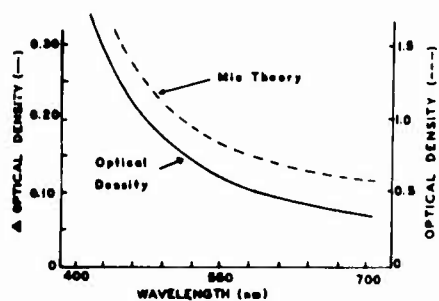


Figure 5. Change in optical density due to ultraviolet exposure and calculated optical density (Mie theory) versus wavelength.

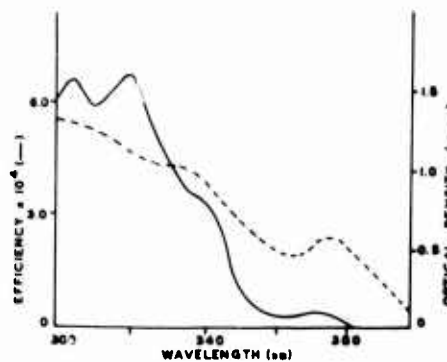


Figure 6. Photodecomposition spectral response compared to optical absorption.

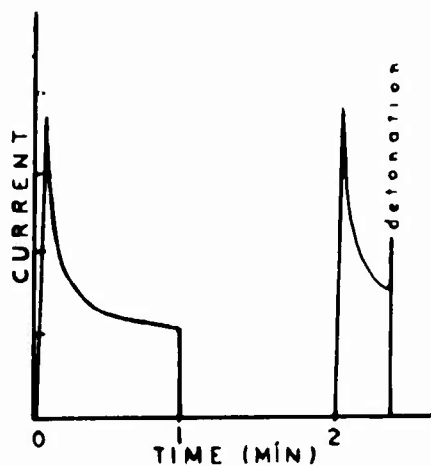


Figure 7. Typical current response to step voltage versus time.

DOWNNS, GARRETT, WIEGAND, GORA
BLAIS, FORSYTH and FAIR

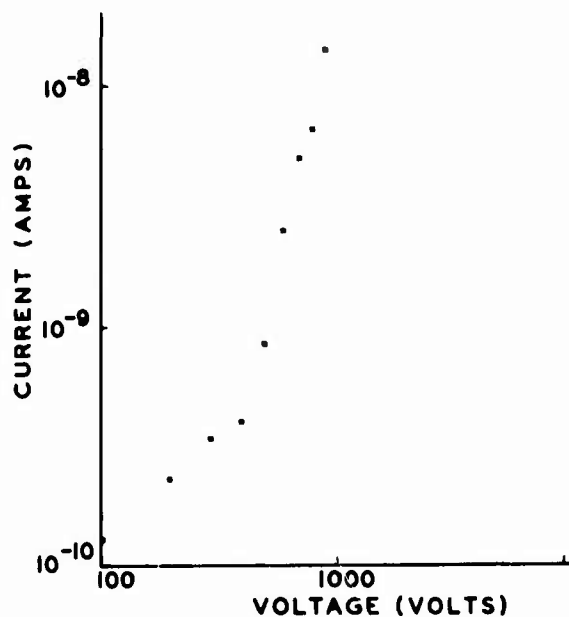


Figure 8. Current-voltage characteristic for .025 cm thick lead azide crystal with gold electrodes.

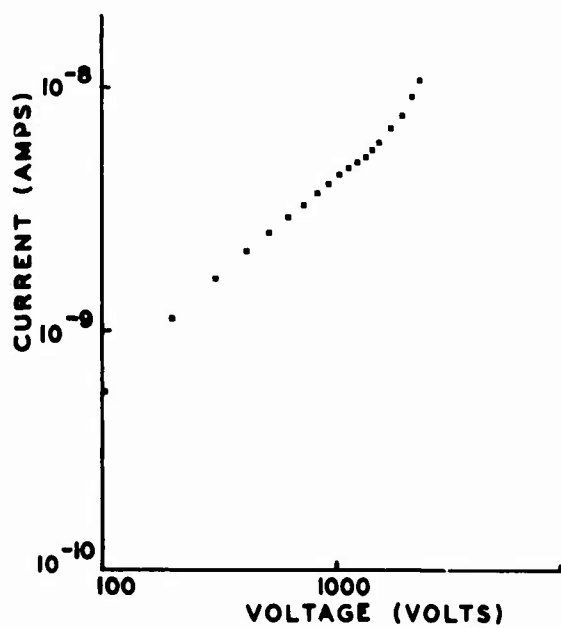


Figure 9. Current-voltage characteristic for .073 cm thick lead azide crystal with gold electrodes.

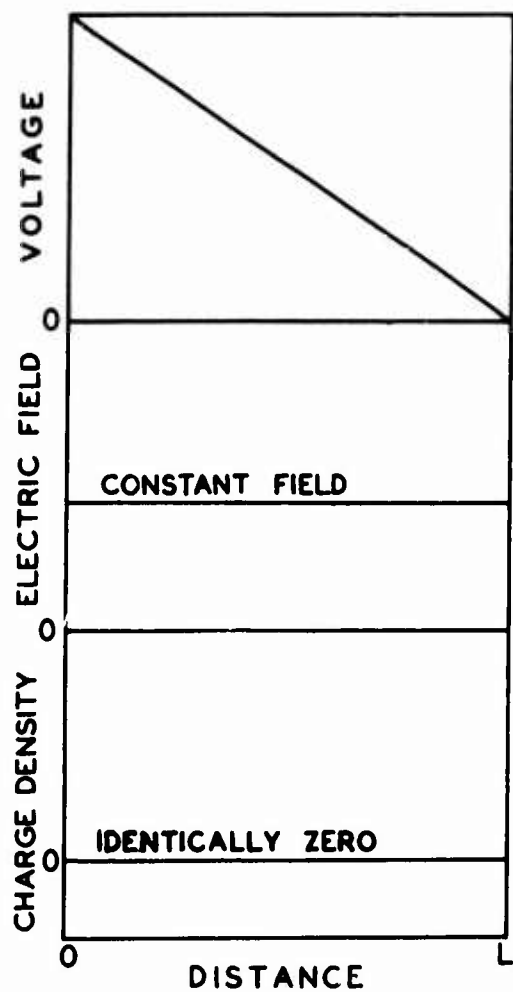


Figure 10. Voltage and charge profiles for the constant-field case.

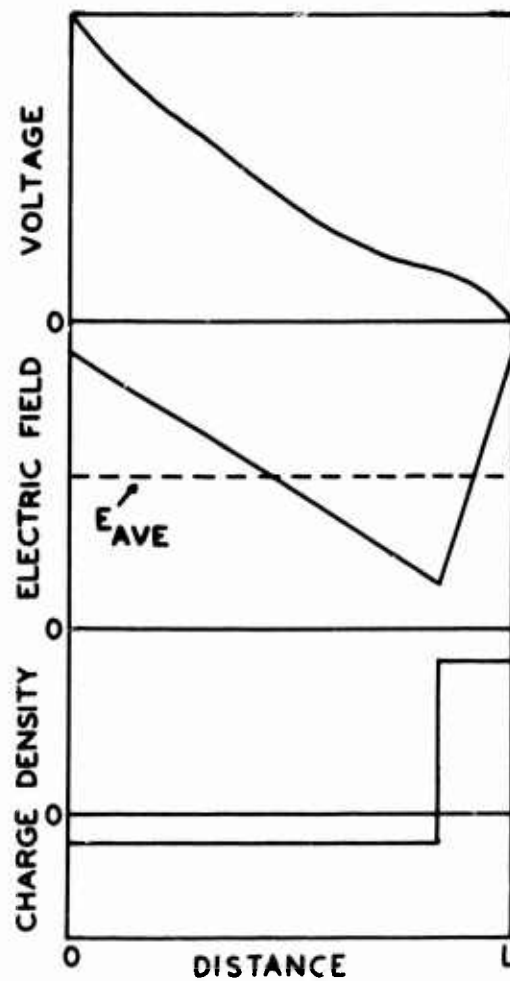


Figure 11. Voltage and field profiles resulting from a non-uniform charge distribution E_{ave} = applied voltage/L.

BOLIVIAN HEMORRHAGIC FEVER IN RHESUS MONKEYS:
TREATMENT WITH SPECIFIC, HOMOLOGOUS ANTIBODY

GERALD A. EDDY, LTC, VC, MICHAEL D. KASTELLO, CPT, VC
STEPHEN K. SCOTT, CPT, VC AND TIMOTHY G. TERRELL, CPT, VC
UNITED STATES ARMY MEDICAL RESEARCH INSTITUTE
OF INFECTIOUS DISEASES, FREDERICK, MD. 21701

To introduce this paper let us consider a hypothetical situation that might develop in a modern military hospital. In this scenario a Captain reports to the dispensary with nonspecific clinical signs including fever, malaise, inappetence and vague aches and pains. The condition is diagnosed as a viral illness, the patient receives appropriate symptomatic medication and is instructed to rest at home and come back in three days if not improved. There is an influenza epidemic at the time and the staff is a bit harassed.

For the great majority of patients there would be no further need for treatment and the dispensary would not see him again, but in our hypothetical situation the patient returns in three days. He is more obviously ill with fever, nausea, occasional diarrhea, severe headache, pain in the lumbar region and shaking chills. He is hospitalized, and a workup by the staff reveals nothing very extraordinary. The leukocyte count is low, the pain in the lumbar area is particularly severe, there is occasional epistaxis and the patient's skin is extremely sensitive. He has several episodes of vomiting and soils the bed linens numerous times. It is of interest also that 16 days prior to entering the hospital he returned from a trip to tropical South America to inspect a Civil Affairs unit that was assisting a disaster relief mission.

Our patient becomes more severely ill, but eventually his fever breaks and the blood pressure drops a bit. He makes a slow uneventful recovery. His discharge diagnosis is that of a severe viral disease of undetermined etiology with complications.

Within two weeks of admission of our Captain there are four more instances of a somewhat similar illness at the dispensary. None are initially recognized as unusual nor can they be distinguished from

Preceding page blank

EDDY, KASTELLO, SCOTT, TERRELL

the many influenza patients. The only clue to something more serious is that one of the new patients is the wife of our Captain and the other three are hospital staff-members. Eventually all find their way into the hospital, and two of them die. One of the senior medical officers recalls some rather similar cases from his service in Korea and the staff is quickly alerted to the possibility of a nosocomial viral infection.

All persons known to have had contact with either the index case or the known secondary cases are contacted and members of the hospital staff who become ill are questioned carefully about contact with patients. Within six weeks from the first signs of illness of the Captain there are at least 17 secondary and tertiary cases and 7 deaths. Each illness has occurred in either hospital employees or in family members of suspected cases. Rumors begin to spread. The disease is thought to be one of the South American hemorrhagic fevers and the appropriate state, national and international health agencies are notified. Virologists, epidemiologists and other specialists arrive en masse along with members of the local and national press. Statements by high government officials are made before nationwide television cameras.

The major local problem at this point however, is fear, particularly in the hospital staff. The virus is identified as Machupo virus, which causes Bolivian hemorrhagic fever (BHF), and the first indications of headache, fever and sore throat in an at-risk staff member raise serious problems. The physician who examines him is in a quandary as to whether to put him on the special contagious ward for hemorrhagic fever patients or to send him home to recover from influenza. Many of the hospital staff refuse to enter the special ward and those who do are fearful for themselves and their families. As the number of cases and deaths mount, panic spreads to the surrounding locality. Fear increases when the disease is diagnosed in a nearby civilian hospital. That diagnosis ultimately proves to be incorrect, but in the ensuing days and weeks day to day interpersonal contacts in the community decrease to an absolute minimum. People who are genuinely ill are afraid to go to the military hospital and the civilian hospitals refuse to admit military personnel or their dependents. Eventually the number of new cases diminishes due to the rigorous precautions initiated in the hospital and dispensary. The disease disappears after having affected 31 people and causing 14 deaths.

This hypothetical situation is not a wild delusion. Recently an outbreak of fatal BHF occurred in a hospital in Cochabamba, Bolivia. But, because the clinical signs and symptoms of the disease were somewhat different from the typical BHF symptoms and because the hospital was well outside the BHF endemic area it was not initially diagnosed. Machupo virus had never been known to exist in Cochabamba.

Eventually it was learned that the index patient had traveled in the Department of Beni, where BHF is endemic, and became ill there before returning home to Cochabamba. This knowledge plus the deaths of the index case and three of the four secondary cases caused a near panic in the city (1). Members of the hospital staff were reluctant to care for a sick pathologist who was exposed during a necropsy of one of the secondary cases and who subsequently became ill and died.

Although other instances of human to human transmission of Machupo virus have been reported (2), this was the first nosocomial outbreak. Hospital associated infections have occurred with the serologically related Lassa fever virus in Africa and the reports of the outbreaks there are remarkably similar to that described for the Cochabamba epidemic (3).

The viruses that caused these illnesses are both members of the arenavirus group which also includes lymphocytic choriomeningitis virus and Junin virus (the cause of Argentine hemorrhagic fever) in addition to a number of avirulent viruses. Some of the more unusual characteristics of infections with the virulent members of the arenavirus group are incubation periods of two weeks or more and prolonged periods of virus shedding (4,5).

Thus, in this era of frequent, rapid international travel the possibility of an outbreak of an arenavirus disease in a large modern hospital becomes a distinct possibility. Our hypothetical situation may never occur, but a realistic assessment suggests that it could. As the remote tropical regions are developed agriculturally and as alterations take place in their ecology, presently known viruses and new viral entities would seem to be candidates for nosocomial outbreaks.

This report describes studies carried out at the United States Army Medical Research Institute of Infectious Diseases during the last 18 months wherein we have defined a rhesus monkey model for studying Machupo virus infections and have successfully treated BHF in monkeys by devising therapy with specific, homologous antibody.

BHF IN RHESUS MONKEYS

In order to work in reasonable safety with Machupo virus our studies were carried out in airtight hoods or in pressurized suits(6). After developing virus plaque assays in Vero cell culture and in vitro tests for virus neutralizing antibody, we assessed the rhesus monkey as a model for this infection by observing clinical signs and by measuring leukocyte counts, viremias, and antibody responses. Complete necropsies were performed on monkeys that died.

Figure 1 shows the observed development of clinical signs in monkeys each inoculated with approximately 1,000 plaque forming units

EDDY, KASTELLO, SCOTT, TERRELL

(PFU) of virus. These signs were similar to those reported in humans (7) except that the onset was earlier in monkeys and the disease course more severe. The earliest monkey deaths occurred about day 15, and 80 to 90% of the monkeys died by day 30, the mean day of death was about day 20.

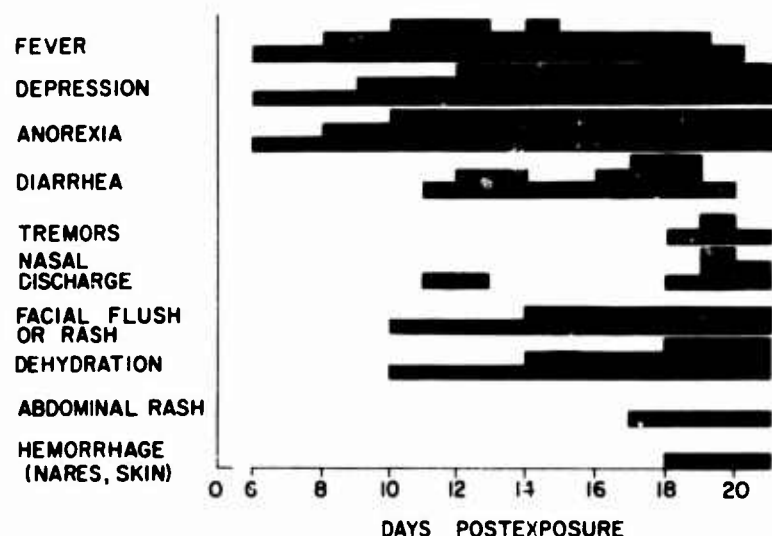


Figure 1. Clinical Signs of BHF in Rhesus Monkeys

Like the reported human cases hemorrhage was not pronounced in the monkeys. There was occasional epistaxis and many monkeys showed some bleeding around the gums. The rash on the face and abdomen did not occur at all in some monkeys but was extremely severe in others.

The neutrophil count dropped precipitously after postexposure day 5 and reached a low on day 13 to 14; other blood parameters that declined included lymphocytes and packed red cell volume. All began to rise on day 13 to 14. Defervescence began at this time although some monkeys had a second fever peak on day 15 to 16.

Histopathology did not reveal consistent hemorrhagic manifestations although hemorrhages were observed in the skin, heart, brain and nerves of some monkeys. Generally, the histopathologic lesions were consistent. There was hepatic necrosis, adrenal cortical necrosis, necrotizing enteritis and epithelial necrosis. Monkeys which died after postexposure day 18 also had nonsuppurative meningoencephalitis and lymphoid necrosis. Most of the histopathologic lesions reported in human autopsies were reproduced in the monkey, however the necrosis of the skin, oral mucosa,

gastrointestinal mucosa and adrenal cortex have not been described in man. Despite these latter discrepancies the rhesus monkey was a satisfactory histopathologic model for the study of BHF.

Ten to 20% of the untreated monkeys in our several experiments survived the initial effects of the virus and began to show some improvement by day 30. They regained their appetite, became more active and were no longer dehydrated, but approximately half of them then developed evidence of central nervous system (CNS) disease. They exhibited paresis, depression, occasional convulsions and incoordination. These clinical signs appeared from day 30 to as late as day 90 and about half of the monkeys died. The histopathologic lesions in the monkeys that died with these signs was primarily vasculitis. Typically it affected all the major organs of the body including the brain.

No neutralizing or complement fixing (CF) antibody was detected in monkeys that died prior to day 21. After that time however, the survivors developed CF titers that gradually rose to 1:512 by day 60. The first neutralizing antibody was detectable by day 28 and reached titers of 1:3200 or more by day 60. These titers are substantially higher than those reported for humans (8) and provided a source of neutralizing antibody for additional studies to define a method of therapy.

The patterns of viremia which developed in monkeys are shown in Table 1. These data represent mean values for several monkeys from three different studies. This relatively prolonged, high viremia is unusual. In most virus infections viremia persists for only 3 to 4 days, and the more severe clinical signs often occur while the viremia is either waning or after it has disappeared. Arenavirus diseases are less acute however, and the prolonged viremia suggests that the pathogenesis might be successfully interrupted with specific antibody therapy. We therefore undertook studies to determine whether convalescent immune serum would protect monkeys if it were administered either at about the time of exposure or after the onset of clinical illness.

PROPHYLAXIS AND TREATMENT OF BHF IN MONKEYS WITH PASSIVE ANTIBODY

Table 2 shows the effect of homologous immune serum given to monkeys that were inoculated 2 hours later with approximately 1,000 PFU of Machupo virus. The quantity of serum given each monkey is shown in the column on the left. The titers on day 4 represent passive antibody and those on day 40 the active antibody titers. The data show that monkeys with low passive antibody titers of 1:4 by CF and 1:8 to 1:32 by neutralization

EDDY, KASTELLO, SCOTT, TERRELL

titers were protected from severe disease. Monkeys with higher passive antibody levels were completely protected from illness but not, apparently, from infection. When the survivors were challenged with Machupo virus on day 60 none became ill. Some monkeys were apparently fully immune despite the absence of detectable neutralizing antibody thereby suggesting a role for cell mediated immunity in this virus infection.

Table 1. Mean viremia titers in Machupo virus infected rhesus monkeys.

Exper. No.	Inoculum Dose	Days Postinoculation (Number Viremic/total)					
		6	10	12	14	17	21
1	3.0*	-	2.7* (5/5)	3.6 (5/5)	4.6 (5/5)	3.7 (3/5)	1.0 (0/1)
2	5.0	2.6 (2/2)	3.4 (2/2)	1.8 (2/2)	-	-	-
3	3.0	-	3.2 (2/2)	2.8 (2/2)	4.8 (2/2)	4.5 (2/2)	2.2 (1/1)

*Log₁₀ PFU of virus

Table 2. Immune serum protection of rhesus monkeys against Machupo virus infection

Serum Dose ml	Reciprocal Day 4				Severity of Illness	Deaths/2	Reciprocal Day 40			
	Antibody Titer						Antibody Titer			
	Neut.		CF				Neut.		CF	
5.0	32	128	16	16	0	0	<8	<8	<4	<4
1.5	32	32	8	8	0	0	8	32	<4	4
0.5	8	32	4	4	++	0	128	512	64	64
0.15	<8	<8	<4	<4	++++	1	2048		512	
None	<8	<8	<4	<4	++++	2	-	-	-	-

Nevertheless, the prophylactic efficacy of antibody was obvious. We therefore assessed the effects of immune serum given to monkeys at various times postinfection. Table 3 shows the results of our efforts to treat monkeys on days 5, 7, 9 or 11 postexposure by giving each monkey 10 ml of immune serum. Bearing in mind that the monkeys became symptomatic on about day 6, it is apparent that the typically fatal disease course was favorably modified after the onset of clinical signs.

Table 3. Treatment of Machupo virus-infected monkeys with homologous immune serum

Day Treated	Severity of Illness	Time to Recovery	Death/2	Reciprocal CF Titer Day 44
5	Trace	1-2 days	0	32 64
7	+	6-8 days	0	64 256
9	+++	2-3 Weeks	0	512 512
11	++++	2 months	1	512 -
None	++++	-	2	- -

In general we were able to reproduce the foregoing results by using ethanol fractionated gamma globulin rather than whole serum. However, treatment after the onset of clinical signs appeared to be less effective with intramuscular injections of gamma globulin than with the equivalent amount of serum given subcutaneously. We reasoned that this may be due to the route of administration and that the deep intramuscular injection was less efficacious than other possible routes.

Table 4 shows the results of giving immune gamma globulin by two routes on day 8 postexposure. The two monkeys treated intravenously suffered a mild clinical disease whereas three of the four monkeys given intramuscular gamma globulin experienced a more severe illness.

EDDY, KASTELLO, SCOTT, TERRELL

Table 4. Effect of route of gamma globulin treatment on clinical illness in Machupo virus infected rhesus monkeys

Route	Severity of Illness (Number affected)	Days to Recovery	Deaths/Total
Intravenous	+(2)	10-15	0/2
Intramuscular	+(1)+++ (3)	10-15	0/4
None	+++ (3)	-	3/3

To quantitate the amount of serum required to treat monkeys early after the onset of clinical signs. Table 5 was prepared as a composite of three studies. It shows the minimum amount of serum or gamma globulin equivalent that resulted in a successful treatment and the maximum amount that did not. In all cases this was given on day 8 and is calculated as the equivalent amount of serum with a neutralizing antibody titer of approximately 1:3000. Thus the amount of potent antiserum required for successful treatment is approximately 3 ml/kg. The human dosage calculated in the column on the right represents the number of 240 ml units of human plasma that would be required to yield similar levels of passive neutralizing antibody in a 70 kg human recipient if the human plasma pool had a neutralizing antibody titer of about 1:100. This represents a substantial amount of plasma, but it could be given as gamma globulin if the need arose.

Table 5. Antibody dosage for early treatment of Machupo virus infected monkeys

Serum Dose (ml/kg)	Severity Illness	Death./total	Equivalent Human Dosage* (Plasma units)
3.0	mild	0/4	26
1.5	severe	1/3	13
None	severe	0/3	-

* See text.

One of the difficulties with antibody therapy later in the clinical course is the rapid turnover that occurs. We observed that monkeys receiving 10 ml of immune serum on postexposure day 10 had barely detectable antibody titers two days later, whereas the data in Table 2 show that monkeys which received 5 ml of serum on day 0 had passive titers of as much as 1:128 on day 4. These data suggested that a single dose of antibody given late in the disease may be removed so rapidly, due to reaction with antigen and to the more rapid turnover of serum protein during severe illness, that successful treatment might require the continuous administration of specific homologous antibody. In a separate experiment virus infected monkeys were treated about 4 days after the onset of illness (day 10 postexposure) with the gamma globulin equivalent of 2 ml/kg of serum intravenously followed by intramuscular doses of one-half that amount daily for the next 10 to 12 days. Although two of the three treated monkeys died, the clinical course and the times to death suggested that such therapy was useful and further studies are now underway.

DISCUSSION

Our data from Machupo virus infected rhesus monkeys showed that species to be a useful model for BHF in humans. The clinical signs, hematologic data and histopathologic lesions were generally similar to those seen in humans except that the disease course was more severe and perhaps somewhat briefer in the monkey. Using the rhesus model we then explored prophylaxis and therapy with homologous antibody. Our data suggest that the use of convalescent plasma might favorably alter an outbreak such as the hypothetical one described in the introduction.

Our data showed that: 1) relatively low prophylactic doses of specific antibody given at about the time of exposure to BHF prevented the development of clinical signs of disease; 2) BHF, unlike most viral diseases, can be successfully treated, in monkeys at least, if antibody therapy is initiated early after the onset of clinical signs; and 3) The maintenance of detectable levels of passive antibody with daily injections may permit successful treatment as much as four days after the clinical signs have become apparent.

On the basis of preliminary data from our BHF studies, the United States Army Medical Research and Development Command funded a project by the Pan American Health Organization (PAHO) and The Middle America Research Unit to collect plasma from BHF immune persons in Bolivia. A number of people in the endemic areas were screened and 16 were selected for plasmapheresis which was carried out by Medical Officers from PAHO and the Bolivian Ministry of Health. A total of 223 units of plasma

EDDY, KASTELLO, SCOTT, TERRELL

was collected and have been sent to the Massachusetts State Laboratory for fractionation of the gamma globulin. The immune gamma globulin will be studied in our laboratory and used in Bolivia.

Although this relatively small amount of antibody will not be adequate for treating large numbers of people, it will provide a means for prophylaxis or therapy of selected medical personnel. There are no BHF immune medical personnel in Bolivia at the present time and it is difficult to persuade susceptible individuals to go into epidemic areas to provide patient care. With a supply of gamma globulin of known potency for prophylaxis and treatment however, the situation of near panic described in Cochabamba may be avoided and medical personnel will be more sanguine about serving in epidemic areas.

In addition to the use of specific human gamma globulin we are exploring the possible use of hyperimmunized monkey serum which may be as much as 100 to 200 times more effective than the available human plasma. This possibility could indeed provide a source of antibody for large scale human therapy if it proves to be safe and effective.

Thus unlike virtually all other viral diseases, arenavirus infections may be unique in that the typical pathogenesis is amenable to specific therapy. Our data suggest that by developing an adequate source, of specific immune serum we may reduce the threat of these rodent-borne viruses.

REFERENCES

1. Peters, C.J.; Kuehne, R.W.; Mercado, R.R.; et al.: Am. J. Epidem. (In press, 1974).
2. Douglas, R.G.; Wiebenga, N.H. and Couch, R.B.: Am. J. Epidem. 82:85-91 (1965).
3. Monath, T.P.; Mertens, P.E.; Patton, R.E.; et al.: Am. J. Trop. Med. Hyg. 22:773-779 (1973).
4. Casals-Ariet, J: Personal Communication.
5. Johnson, K.M.: Am. J. Trop. Med. Hyg. 14:816-818 (1965).
6. Kuehne, R.W.: Appl. Microbiol. 26:239-243 (1973).
7. Mackenzie, R.B.; Webb, P.A. and Johnson, K.M.: Am. J. Trop. Med. Hyg. 14:1079-1084 (1965).

AUTOMATIC SHELL FRAGMENT
MEASURING TECHNIQUE

JAMES W. FASIG
MATERIEL TESTING DIRECTORATE
ABERDEEN PROVING GROUND, MARYLAND

Good afternoon gentlemen, I'm Jim Fasig from the Materiel Testing Directorate of Aberdeen Proving Ground. It is a pleasure to have this opportunity to describe to you one of the many tests performed on explosive devices, to relate some of the analysis procedures, and to describe an instrument my group developed for improving this analysis process. This discussion is unclassified.

First, I will briefly describe the procedures and facilities used in testing explosive projectiles so you will be able to relate to the total test program.

When weapon systems which utilize high explosive warheads are being developed or redesigned, one major question is - how effective is the new warhead? High explosive munitions are primarily designed to produce lethal effects through fragmentation. Therefore, effectiveness associated with high-explosive warheads is a measure of the ability of the fragments produced to inflict casualties on enemy personnel in the area of detonation.

Now it must be remembered that warhead effectiveness is only one part of the total effectiveness analysis of the weapon system - accuracy, reliability, safety, mobility, vulnerability and other factors must be evaluated to analyze the expected system performance. The determination of warhead performance is the measure of effectiveness of the weapon system where the payload or warhead is on target. Criteria will be stipulated for this performance and inability to meet requirements will result in either redesign and or warhead rejection.

The most widely used and accepted measure of warhead effectiveness is lethal area. Lethal area is a calculated term with the dimensions of area. It is used by weapon system analysts to estimate the number of casualties that would be caused under a specific set of firing conditions. Briefly, the expected number of casualties is defined mathematically by the expression,

$$E_c = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma(X,Y) P_K(X,Y) dx dy$$

where $\sigma(X,Y)$ is the density of troops at a specific point (X,Y) on the ground, and $P_K(X,Y)$ is the probability that a soldier located in the elemental area $dx dy$ will be incapacitated. On the assumption that the targets are uniformly distributed over the ground plane, $\sigma(X,Y)$ will be a constant and the expression can be written

$$\frac{E_c}{\sigma} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} P_K(X,Y) dx dy$$

The quantity E_c/σ has the dimensions of area and is defined as lethal area.

It is apparent that the incapacitation probability is a function of the total fragmentation characteristics of the munition being considered. More precisely it is a function of the weight, velocity, number and shape of the fragments produced. To obtain the best estimate of a warhead's lethal area we would need to collect the weight, velocity and shape for each fragment produced by the test munition. Since a large warhead will produce 25,000 fragments this approach is impractical. Instead fragment samples are obtained in an arena test area.

The arena (Figure 1) is divided into two parts - one side is used to measure the fragment velocities. You will note that the sample areas are shown on the panels (Figure 2). These are radial zones projected from the axis of the projectile. The projectile is placed in the arena center and detonated. The fragment velocities are obtained photometrically and produce a spectacular display on the velocity panels. Since the fragments exceed the velocity of sound they arrive at the panels before the blast wave destroys the arena

(Figure 4). The fragment velocities (Figure 5) from a typical firing are highest at the nose, 90° from the nose, and at the base of the shell. The other side of the arena is used to recover shell fragments. About 10% of the fragments produced are captured. All the captured fragments are counted and weighed. Based on the weight distribution a sample of fragments is selected for measurement of average fragment area. Since fragments are captured in zones the density of fragments is also obtained. Here again, the density of fragments about the projectile is greatest perpendicular to the shell axis (Figure 6).

A shape factor, determined for a sample of the captured fragments, is related to the fragment's weight and its average presented area. The presented area is the plane area confined within the outline of the fragment in a particular orientation. The shape factor is required for two reasons: first, the fragment striking velocity is functionally dependent on the shape factor; second, penetration limits for armor and personnel are defined based on this factor.

The procedure for determining the shape factor is to randomly select a sample of fragments, determine for each one its average presented area, and then by least squares establish the functional relationship between fragment weight and its average presented area.

The average presented area for a fragment is determined by measuring its presented areas in 16 equally spaced orientations, and computing the arithmetic mean of the individual observations. The 16 orientations are specified along axes obtained by striking lines from the center to the 10 faces and 6 vertices which constitute one-half of a regular icosahedron (Figure 8). The regular icosahedron is a 20-sided regular polyhedron, each side formed by an equilateral triangle of the same size.

The measurement technique used in the past employed a micro-projector and a pivoted slip-ring assembly (Figure 9). The assembly was designed such that by pivoting the table containing the slip-ring and rotating the slip-ring the angles defining the 16 required orientations could be obtained. With a fragment on the cross-hairs at the focus of the lens system, a shadow image was projected on the ground glass screen of the micro-projector and hand traced. The presented areas were determined by using a planimeter around the traced area. This technique was tedious and time consuming. It rarely allowed more than a 30-fragment sample to be measured per test. To increase efficiency and to meet revised fragmentation data requirements for at least 100 samples per test round, my group developed an instrument which automated the measurement of presented area.

FASIG

The instrument can be best understood if divided into its two separate functions - measurement of presented area and the automatic control circuitry.

The measurement system consists of four identical collimated light systems and a rotating ring with crossing wires on which the fragment is mounted. The four collimated pairs are positioned to determine the area of the shell fragment in the 16 equally spaced orientations of the icosahedron.

The measurement principle is very simple and is based upon a system described in most physics textbooks. When the fragment is placed in a collimated light source the shadow formed is identical to the projected area of the fragment. Collimated light is needed, of course, to provide sharp shadow contrast at fragment edges and to preclude highlights from shiny parts of the fragment.

Each collimated pair consists of a transmitter and receiver section. The collimated light of the transmitter is produced by using a light emitting diode (LED) at the focus of a 24-inch lens system. The LED is modulated so that background light in the room is eliminated in the measurement. A uniform light beam of about 3-1/2 inches in diameter is obtained. Because fragments with lengths from 3" to 0.1 inch are to be measured a shutter system is provided having 3 different sized apertures. The aperture regulates the area of the collimated light beam transmitted and thereby reduces the required measurement circuit range when a small fragment is being measured.

The receiver section consists of another lens system with a photo diode at the focal point. The photo diode provides an output voltage proportional to the amount of the light focused on it. This voltage is gated to an analog-to-digital converter where it is subtracted from the total aperture area, thereby producing a value equal to the area of the fragment. This value is both printed and stored. The stored areas are combined to obtain the average presented area after all 16 measurements are completed.

The control circuitry consists of a system to automatically measure the areas presented in the 16 separate views by switching collimator pairs and taking the five required measurements for each of three tilted collimators, and the single measurement from the vertical pair. The circuit is strobed by magnetic pickups along the periphery of the index ring.

A second circuit is used to remotely position the shutters for the various fragment sizes. Over 200 integrated circuits are

FASIG

used to control the measurement and indexing process, leaving the operator with only fragment mounting and press to measure (Figure 10).

The instrument has been on line for nearly 2 years. It has required little maintenance. Fragment measurement times have been reduced to 50 seconds versus the 30 minutes previously required, thus providing the needed increase in sample size for the establishment of shell fragment shape factors.

The ability to determine the presented areas of a large number of fragments leads directly to a desire to increase efficiency in the recovery and weighing of fragments. Therefore, an automatic weighing system has been obtained and a new method for fragment recovery is being studied. This concerted effort to reduce test time and cost in establishing munition lethality characteristics is essential if weapon system analysts are to have the information necessary to select effective weapons for further development and to maintain quality during production.

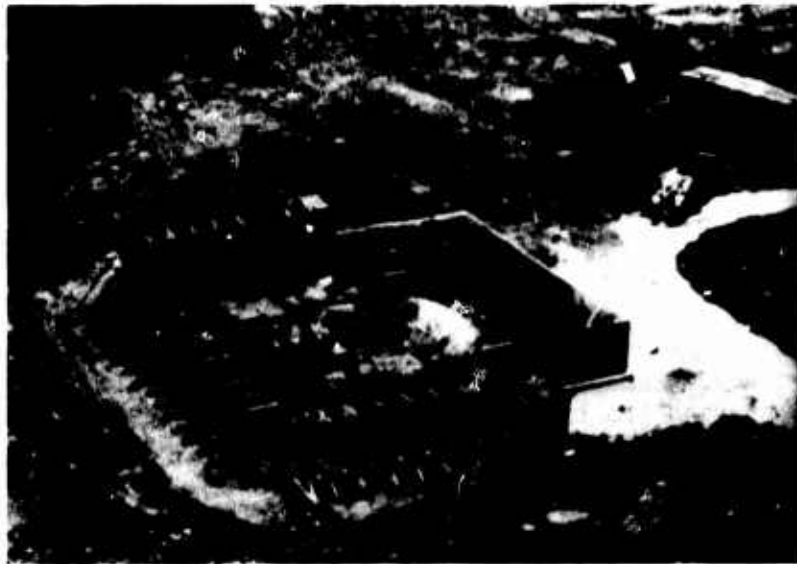


FIGURE 1. FRAGMENTATION TEST AREA



FIGURE 2. FRAGMENT IMPACT GRIDS

FASIG

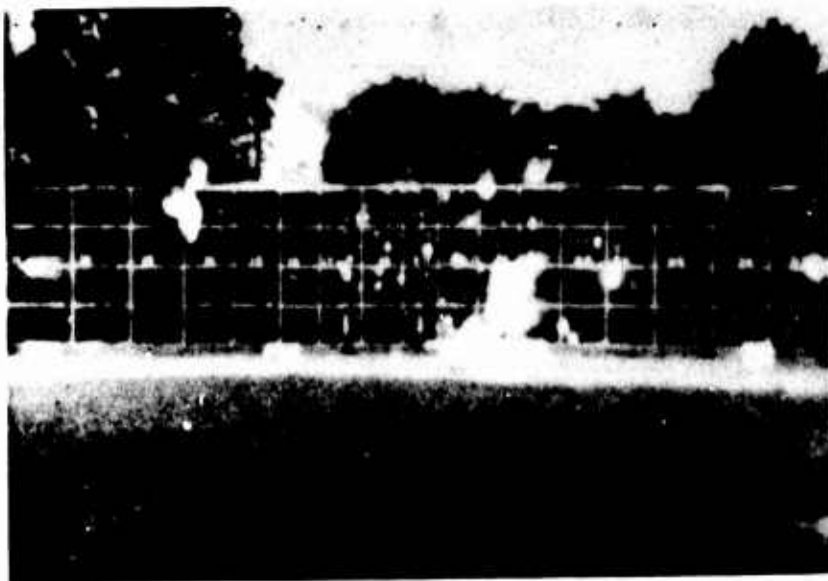


FIGURE 3. FRAGMENT IMPACTS ON GRIDS

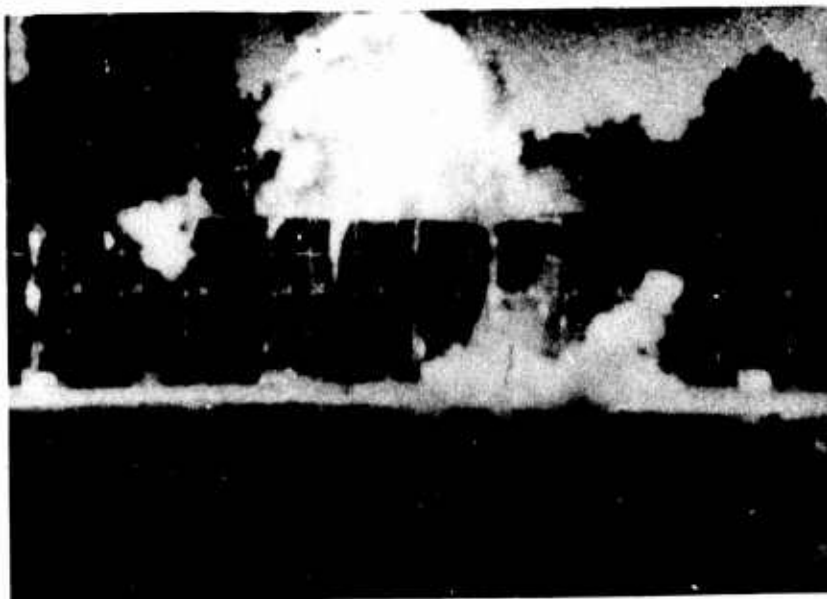


FIGURE 4. WARHEAD DETONATION IN TEST ARENA

FASIG

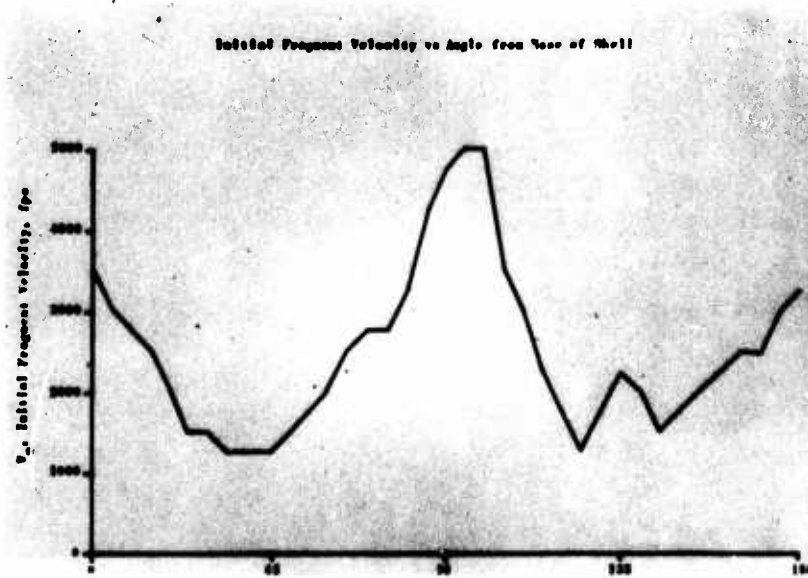


FIGURE 5. INITIAL FRAGMENT VELOCITY VS
ANGLE FROM NOSE OF SHELL



FIGURE 6. FRAGMENT SPRAY DENSITY VS
ANGLE FROM NOSE OF SHELL

FASIG

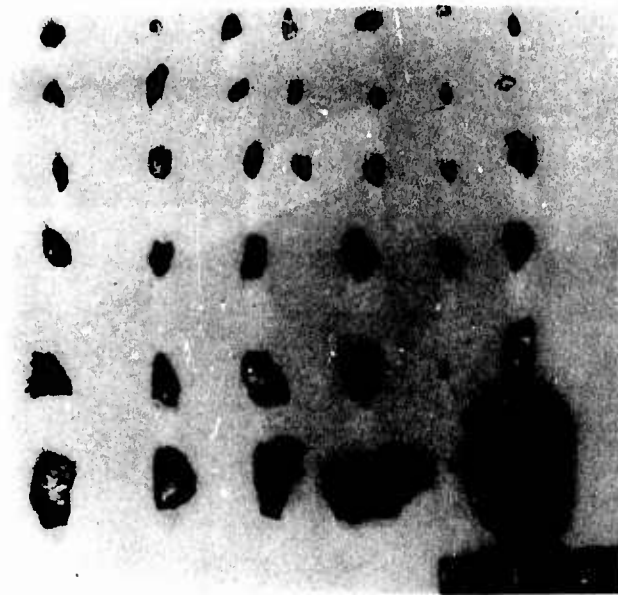


FIGURE 7. SAMPLE OF RECOVERED FRAGMENTS

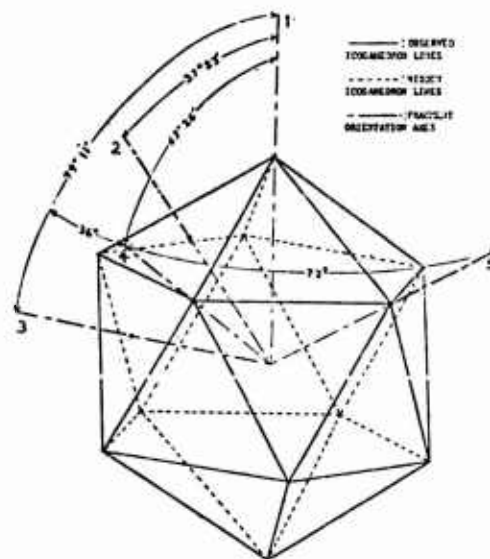


FIGURE 8. MEASUREMENT ANGLES GENERATED USING ICOSAHEDRON

FASIG



FIGURE 9. ONE OF SIXTEEN AREAS TO BE TRACED
USING MICRO-PROJECTOR



FIGURE 10. AUTOMATED SHELL FRAGMENT AREA
MEASURING SYSTEM (ASFAM)

RAPID CASE DESIGN
USING FINITE ELEMENT STRESS ANALYSIS AND NASTRAN

DIANA L. FREDERICK AND JOHN E. HIGGINS
FRANKFORD ARSENAL
PHILADELPHIA, PENNSYLVANIA

INTRODUCTION: The finite element approach to structural analysis came into being with large scale scientific computers. Implementation of the finite element approach to structural analysis lagged behind the theory until the release of CRTPLS and NASTRAN in 1970. The release of these programs provided a substantial analytical capability at a low cost. Although there are many technical improvements to come, the primary challenge is the effective use of existing capabilities and efficient design of structures. This paper concerns this challenge and how it applies to product engineering of ammunition components.

The Finite Element stress analysis technique has been programmed to obtain optimum cartridge case designs for ammunition for personnel weapons and heavy machine guns where formerly empirical or classical methods were used. For example, in conjunction with feasibility studies of cartridge cases, Frankford Arsenal developed computer program(s) capability to improve the technology in ammunition cartridge case design.

GRIDXY (Mesh Generator): Finite Element programs require large quantities of input data to describe the geometry, material properties, boundary constraints and driving functions of the body to be analyzed. Manual preparation of this input requires considerable time (70 to 100 manhours) and the accuracy of the resulting input is questionable. The program, GRIDXY, automates the generation of the Finite Element input, reduces the time required to 2 to 3 manhours and increases the accuracy of the input.

Preceding page blank

FREDERICK AND HIGGINS

Input for the grid generator program consists of describing the geometry of the body to be analyzed and the desired grid spacing, material properties, and boundary conditions. The program generates the grid, numbers the nodes and elements in the proper sequence, and applies the material and boundary codes to each of the elements, as shown in Figure I. The output of this program can, with the proper data manipulation be used as input to the NASTRAN or CRTPLS Finite Element programs.

CRTPLS: CRTPLS is a finite element program developed to analyze cartridge cases. It is sufficiently general to analyze any axisymmetric or plane structure undergoing elastic-plastic quasi-static loading. The approach used in the program is the incremental loading method, wherein at each incremental load step a new stiffness matrix is formulated and solved for incremental deformations, stresses, and strains.

The stresses and strains for the desired load steps are saved on magnetic tape. The stresses saved are the principal stresses, hoop, axial, radial, and shear. The strains saved are the total hoop, axial, radial, and shear strains, and plastic strains. The Von Mises stress is also calculated but not saved. The saved output is usually represented in a plot, of stresses and deformations, Figure II.

NASTRAN: Unfortunately, the CRTPLS stress analysis program for optimum case design cannot be used for other ammunition component structural analyses problems or for even some aspects of cartridge case/diameter interface interactions. We then turned to NASTRAN.

NASTRAN, a general purpose computer program developed by the National Aeronautics and Space Administration, is designed to analyze the behavior of elastic structures under a range of loading conditions, using a Finite Element displacement method of approach. The program is applicable to almost any type of linear and some nonlinear structures that can be represented by combinations of elements contained in the NASTRAN library, such as beams, rods, shear and twist panels, triangular and quadrilateral plates, conical and toroidal shells, solids of revolution, scalar elements, general elements, and constraint elements.

A wide range of analytic capability has been built into NASTRAN, including static response to concentrated and distributed loads, to thermal expansion, and enforced deformation; dynamic response to transient loads, to steady-state sinusoidal loads, and to random

excitation; determination of real and complex eigenvalues for use in vibration analysis, dynamic stability analysis, and elastic stability analysis. In addition, there is a limited capability for solving nonlinear problems, including piece-wise linear analysis of nonlinear static response and transient analysis of nonlinear dynamic response.

The nonlinearity of a structural element is defined by the material characteristics of the elements. Any isotropic material may be made nonlinear by including a stress-strain table, defining its elongation characteristics. The stress-strain table must define a nondecreasing sequence of both stresses and strains, because the stiffness matrix for the first slope should correspond to the defined modulus of elasticity. Linear elements and materials may be used in a more efficient manner than the nonlinear elements since there are no extrapolations. The nonlinear effects depend on the element type. The elements which utilized the plastic material properties are ROD, TUBE, BAR and PLATE elements.

DISCUSSION: A useful and practical exercise of nonlinear analysis in the field of ammunition structural analysis, using RING elements, is the interaction between the cartridge case and gun barrel chamber. In particular, the model enables the effects of changing dimensional tolerances, material mechanical properties, and geometrical variations on the distribution of stresses to be conveniently estimated.

A characteristic problem is a case neck separation malfunction. Of specific interest is the rupture or separation which occurs at the neck-shoulder section. The rupture is circumferential, occurs long before peak pressures are attained, and is closely related to the ammunition/weapon interface. The basic analytical problem concerned tracking the case elements in the stress and strain domain as each increment of pressure is added. Tracking covered taking up initial clearances and continuation with support by the chamber. Rupture of the type being studied would be expected during the complicated interface interactions.

As background, case neck separations (CNS) have occurred in the M61A1 (6 barrel) or M197 (3 barrel) type Gatling guns. No CNS have been reported from firings in the M39 type revolver gun; no CNS have been reported during case and cartridge acceptance firing tests; no CNS have been reported with the firing of M56 (HEI) ammunition; all CNS have occurred with M55 (TP) and M220 (TPT) cartridges.

Case neck separations present serious problems in double ended linkless feed systems, i.e., systems in which the fired cases are

returned to the storage drum. Round control of cases with partially or completely separated necks is lost by the hand-off sprockets, resulting in a system jam. In aircraft systems where the fired cases are dumped overboard, case neck separations present no problem unless the separated case neck remains in the gun barrel chamber. This can result in a jam when a subsequent round is fed into the same barrel.

Case neck separations were initially reported in early A7D aircraft, which have an internally mounted M61A1 gun with a double ended linkless feed system. Results of the malfunction investigation indicated that case neck separations were related to the neck-shoulder blend radius of the chamber and cartridge case. The NASTRAN model exercise is designed to include both material properties and geometry variations.

When a round is fired, the propellant pressure builds up and the sidewall of the cartridge case expands elastically to its yield point and then completes its expansion plastically. Although the sidewall may or may not enter the plastic range before taking up the initial clearance between the case and the chamber, it will be completely plastic when the pressure reaches the maximum value. At this instant of maximum pressure, both the outside diameter of the case and inside diameter of the chamber will have expanded together to a common maximum value. Here the cartridge case sidewall will be acted upon by the internal pressure on the inside by the chamber-cartridge case interface friction and pressure on the outside. The chamber wall will be acted upon by equal and opposite friction and pressure. Knowing the radial and axial loads on the cartridge case at this instant of maximum pressure, the associated state of stress in the sidewall can be determined by applying the Von Mises law of yielding, together with its associated flow rule.

In the problem of wall expansion of the cartridge case and barrel chamber which is caused by the pressure of propellant gases it is desired that the axisymmetric solid of revolution RING element be utilized. This element offers both simplicity and accuracy over other elements. NASTRAN requires augmentation to accommodate plastic analysis of RING elements. Based on the results obtained using the RING element, the manual piecewise linear analysis appears to give accurate results.

ASPECT RATIO ANALYSIS: During the course of the investigation it was realized that the aspect ratio would be very large for the finite elements to be used in the NASTRAN model of the cartridge case. The aspect ratio is defined as the ratio of the element

length to its height. From past experience in using NASTRAN and from consultation with NASTRAN experts, it was more or less understood that the aspect ratio for PLATE elements should not exceed three, and this is most likely true for the RING elements. However, since RING elements are seldom used, no basic study on the applicability of the RING elements as a function of aspect ratio has been made; therefore, a study was required to determine if the problem of aspect ratio in RING elements could be ignored.

A list of all the cases investigated for various aspect ratios and the total number of elements is shown in Figure III. There were two models used in this study, a thick wall steel cylinder and a thin wall brass cylinder. For both studies the cylinder size remained constant but the element size varied giving a wide range of aspect ratios and model variation.

The finite elements used in synthesizing the NASTRAN model of the thick wall cylinder is shown in Figure V. The tube was assumed to be free at both ends, with internal pressure applied. NASTRAN rigid Format 1 and RING elements were used. The overall model had 85 RINGS (or grid circles) and 64 elements for a total of 151 degrees of freedom. The material for the model was steel, with a Poisson's ratio of 0.3. A bilinear stress-strain curve was selected for the elastic-plastic material property.

A total of 17 load increments was made manually for the 16-layer tube. This is the minimum number of load increments since each load increment was pre-calculated to make the material in the subsequent layer plastic. In this problem, elements along the same layer were assumed to behave the same. Although this isn't necessarily true in actual problems, it can still be handled by checking each element and adding more increments.

The increment displacements and stresses were cataloged and filed, then added to the previous results to obtain the total displacements and stresses. After each load increment, the stresses were tested with the Tresca yield conditions. The elastic material properties of those elements that satisfied the yield criterion were changed into plastic material properties.

In order to gain additional confidence on the manual piece-wise linear approach, a thin wall cylinder was considered since the wall of a cartridge case is very thin with respect to its diameter. The finite elements used in synthesizing the NASTRAN model of the thin wall cylinder are shown in Figure IV. The shell was assumed to be free at both ends, with internal pressure applied. Rigid Format 1

and RING elements were used. The overall model had 25 RINGS (or grid circles) and 16 elements, for a total of 45 degrees of freedom. The material selected for this model was brass, with a Poisson's ratio of 0.3 in the elastic range and 0.45 in the plastic range. A bilinear stress-strain curve was selected for the elastic-plastic material property. Only five increments were needed since there were only four layers through the thickness of the wall. The same procedures were followed as in the thick wall cylinder case.

The significant results from the Aspect Ratio Study are presented in Figure V. The results from the NASTRAN analysis for displacements and radial and circumferential stresses are compared with the theoretical results, using the classical equations for the solution of elastic analysis in thick and thin walled cylinders. This figure depicts radial stress (psi) versus wall thickness (inches). The circles are data points for the NASTRAN run with an aspect ratio of 1:100 with 32 elements. The triangles are data points for the NASTRAN run with an aspect ratio of 1:50 with 16 elements. The solid black line represents the thick wall cylinder theory. The applied force is 20,000 lbs. or 21,220 psi.

The figure shows that reasonable agreement between the closed form solution and the NASTRAN model was obtained and thus the aspect ratio can safely be ignored.

ELEMENT SIZE STUDY: In modeling the cartridge case, the elements around the neck are small in size. If all the elements are kept the same size, the number of elements in the cartridge case model is very large. In turn, the run time for the job is long, and thus, expensive. The number of elements is reduced by increasing the aspect ratio of the elements other than those in the cartridge case neck.

A study was initiated with cylinder models shown in Figure VI. The tubes were assumed to be free at both ends, with internal pressure applied. The material of the thin wall model was brass, with a Poisson's ratio of 0.3. Only elastic analysis was considered.

The results for displacement, and radial and circumferential stresses from the NASTRAN analysis are compared with the theoretical results, using the classic equations for the solution of elastic analysis in cylindrical shells. Similarly the results show that we can decrease the number of elements in our model by increasing the element size in the areas other than those in the cartridge case neck.

FREDERICK AND HIGGINS

MODELING TECHNIQUE: Figure VII is the model used for the case neck taper study. The lower model is the cartridge case and the upper model is the chamber wall. With a minimum dimension cartridge case and maximum dimension chamber, there is a small area allowed for the cartridge case to expand unconstrained. The arrows represent forces applied on the cartridge case. The initial NASTRAN run exclude the chamber elements to save time and cost. When the displacement reaches the maximum allowed, the chamber elements are added. In addition the NASTRAN model allows three stress strain curves to be included.

A step by step procedure is described here as the propellant pressure increases and the sidewall expands when a round is fired.

1. The incremental loads are applied to the NASTRAN program to obtain a solution. Every element and grid point is checked after each run.
2. The sidewall expands elastically to its yield condition. This is done by applying the Von Mises law of yielding, together with its associated flow rule. The material properties of those elements which exceeded the yield point are changed according to the appropriate stress strain curves. This continues until the entire sidewall of the cartridge case becomes plastic (change in the slope of the stress-strain curve). Although the sidewall may or may not become plastic before taking up the initial clearance between the case and the chamber, it will be completely plastic when the pressure reaches its maximum value. The fracture point, also, might be reached before the maximum pressure is reached.
3. Check to see if the sidewall of the cartridge case touches the chamber. The displacement of the outer wall of the cartridge case is observed very closely for every element. When the value of the displacement of any point reaches the value of the clearance between the case and the chamber, the points are connected. This will continue until all points on the case are supported by the chamber. From here on the outside diameter of the case and inside diameter of the chamber will expand together to a common maximum value at the instant of maximum pressure. Here the cartridge case sidewall is acted upon by the interface friction and pressure.

In conclusion NASTRAN is an operational program. It can be procured for a few thousand dollars with little or no costly or time-consuming software development for basic analysis capability. The program is general - given the right type of element almost any structure can be analyzed with complete generality in material

properties. The program is adaptable - various stages in the design process can be modeled, crude models early in the process, very detailed in later stages. The program is divisible - substructures can be modeled and later combined into a whole.

NASTRAN cannot design the component - an idea is needed. It cannot predict fatigue life - time is needed. It cannot assure manufacturing feasibility - a process is needed. It cannot assure packaging feasibility - a human is needed.

In summary:

1. Elastic-plastic problems with static loading can be solved using the manual piece-wise linear analysis.
2. The manual piece-wise linear analysis is more accurate than the automated piece-wise linear analysis, NASTRAN Rigid Format 6.
3. RING elements, which are convenient to use for axisymmetric bodies, appear to give accurate results.
4. Problems of aspect ratio that existed in the PLATE elements, this can be safely ignored with respect to RING elements.

REFERENCES:

1. Yang, J.C.S., and Frederick, D., Application of Nonlinear Analysis (Plastic) to NASTRAN Using RING Elements Including Aspect Ratio Effects, Frankford Arsenal Report TN-1178, August 1973.
2. Brophy, J.M., Computer Aided Cartridge Case Design Using Finite Element Stress Analysis: The Automation of Finite Element Configuration, Frankford Arsenal Report R-2054, September 1972.
3. The NASTRAN User's Manual, NASA SP-222, Section 3, September 1970.
4. The NASTRAN Theoretical Manual, NASA SP-221, Section 3, September 1970.
5. NASTRAN Demonstration Problem Manual, NASA SP-224, pp6. 1-1 through 6, 1-12, September 1970.
6. Fishman, H.M., Reddi, M.M., and Zudans, Z., User's Manual for the CRTPLS Computer Program, Franklin Institute Research Laboratory Report F-C1865-05, September 1970.

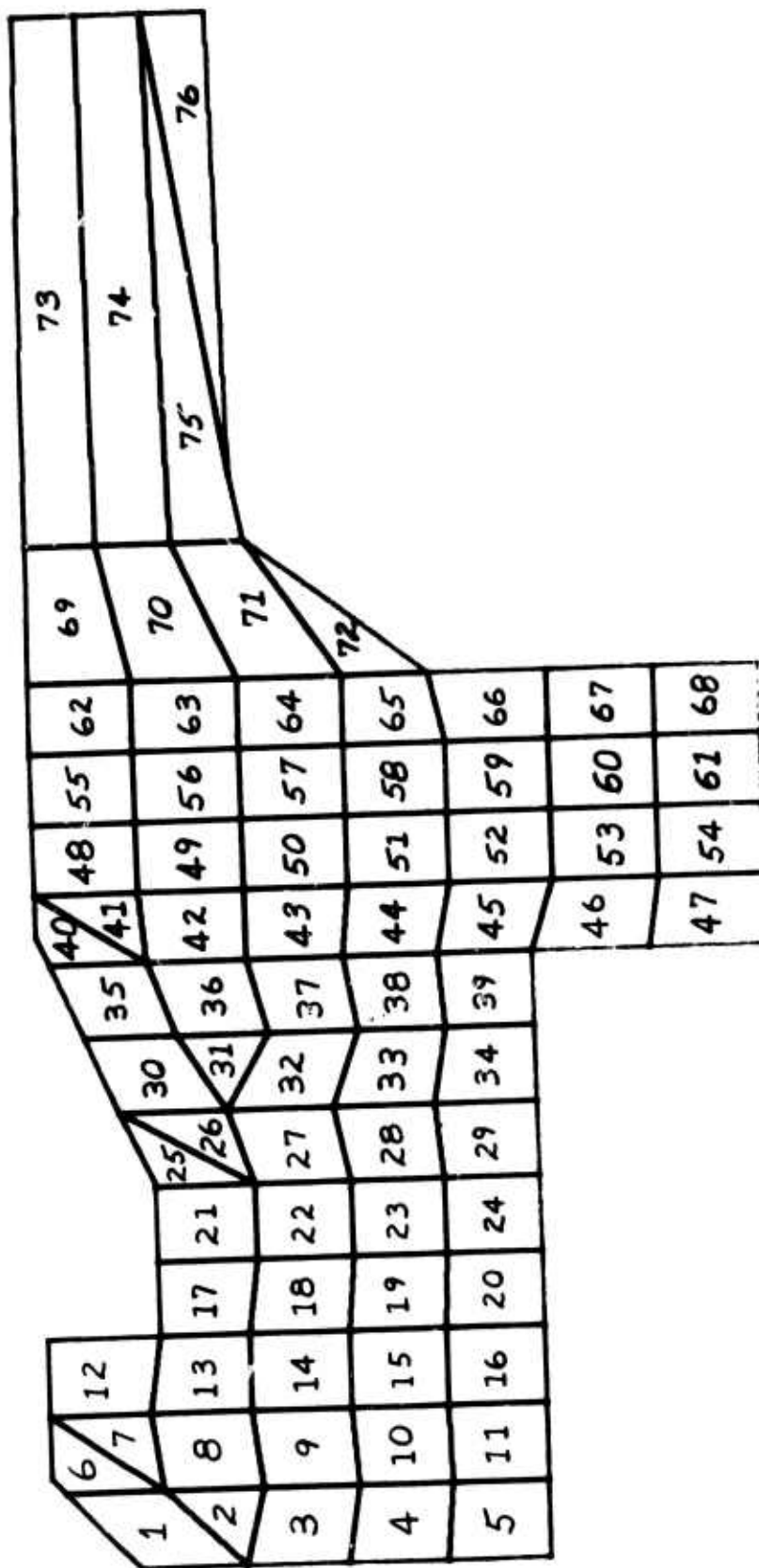


FIGURE 1. GRIDXY PLOT

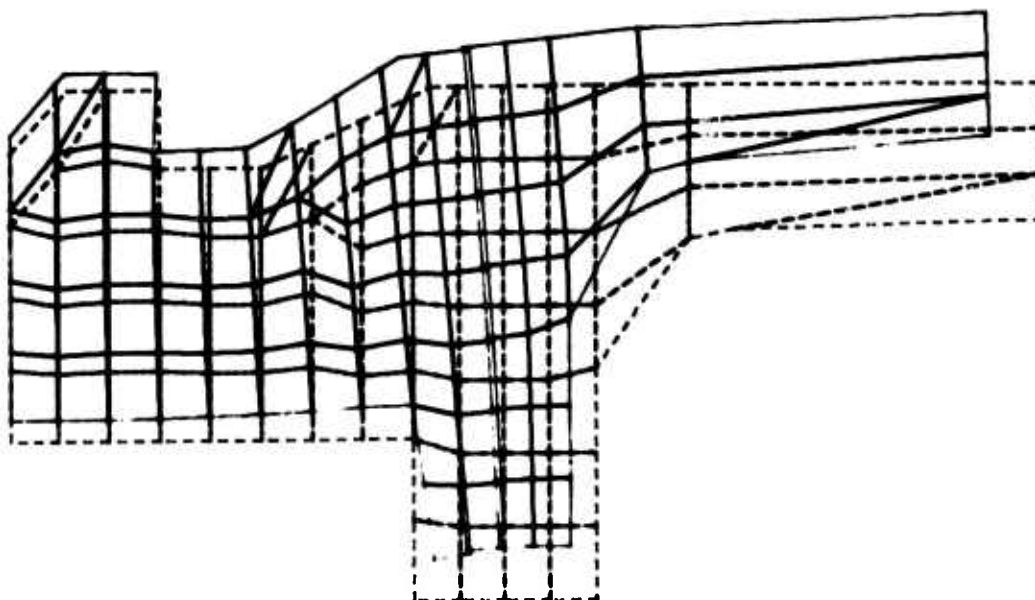


FIGURE IIa. DEFORMED GEOMETRY PLOT

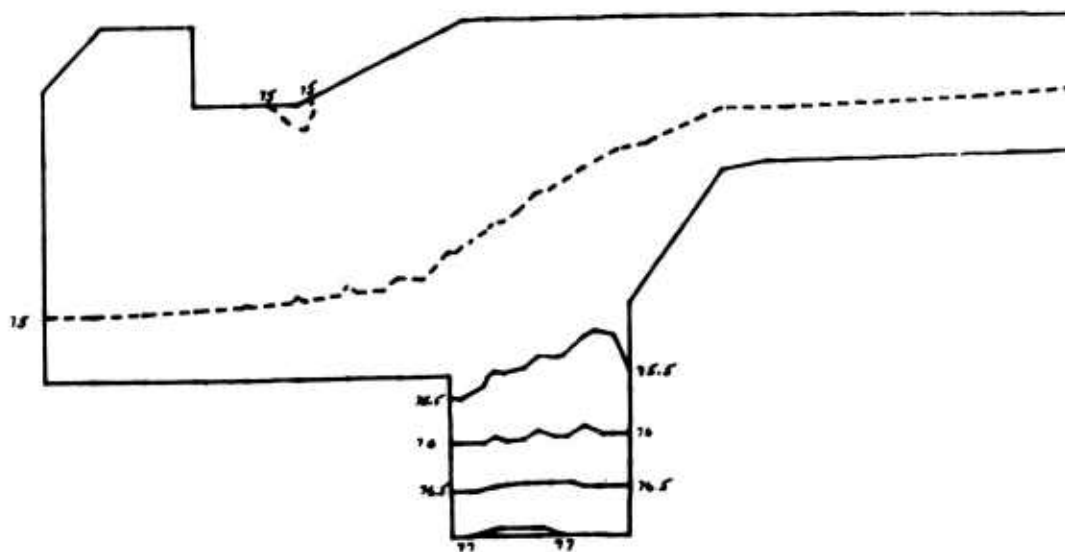


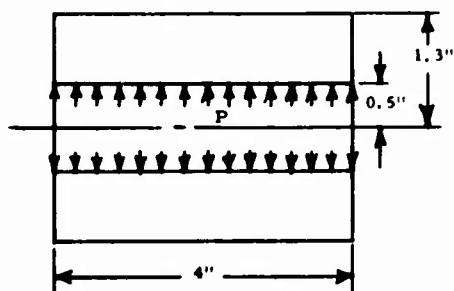
FIGURE IIb. ISOSTRESS PLOT

<u>MATERIAL</u>	<u>RATIO</u>	<u>ELEMENTS</u>
STEEL	2:5	32
	1:5	64
	1:10	128
	1:20	64
	1:10	32
	1:5	16
	1:40	128
	1:80	256
BRASS	1:160	128
	1:1	64
	1:50	32
	1:100	32

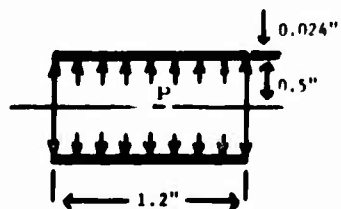
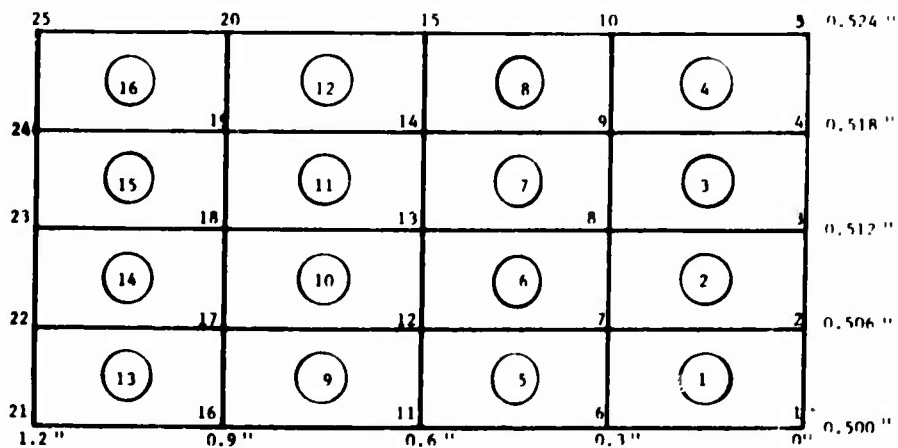
FIGURE III
ASPECT RATIO STUDY

FREDERICK AND HIGGINS

85	68	51	34	17	1.3 "
84	67	50	33	16	1.25 "
83	66	49	32	15	1.2 "
82	65	48	31	14	1.15 "
81	64	47	30	13	1.1 "
80	63	46	29	12	1.05 "
79	62	45	28	11	1.0 "
78	61	44	27	10	.95 "
77	60	43	26	9	.9 "
76	59	42	25	8	.85 "
75	58	41	24	7	.8 "
74	57	40	23	6	.75 "
73	56	39	22	5	.7 "
72	55	38	21	4	.65 "
71	54	37	20	3	.6 "
70	53	36	19	2	.55 "
69	52	35	18	1	.5 "
4 "	3 "	2 "	1 "	0 "	

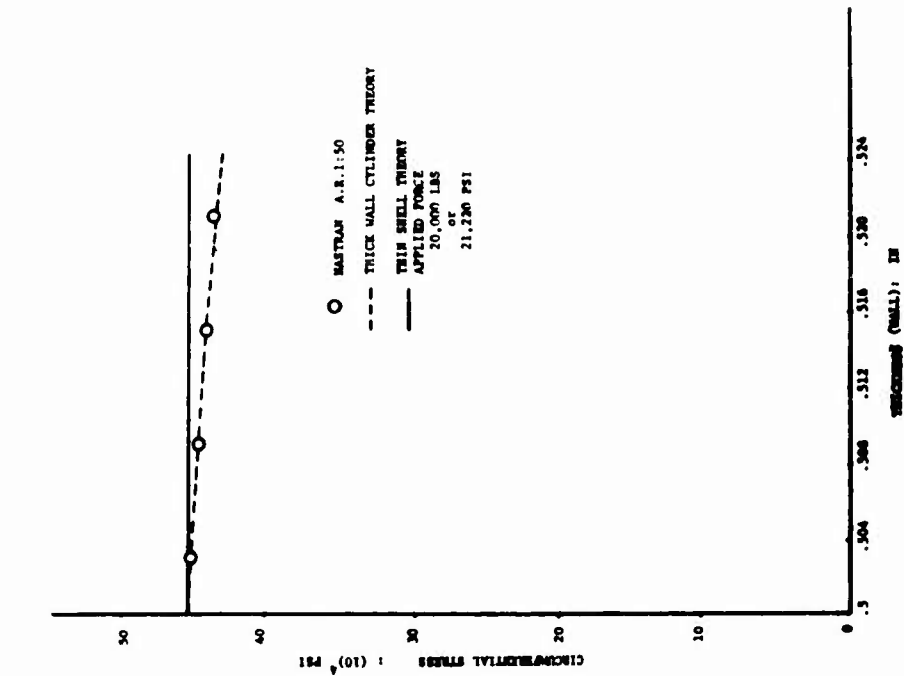


NASTRAN Model, Thick Wall Cylinder. Piece-Wise Linear Analysis

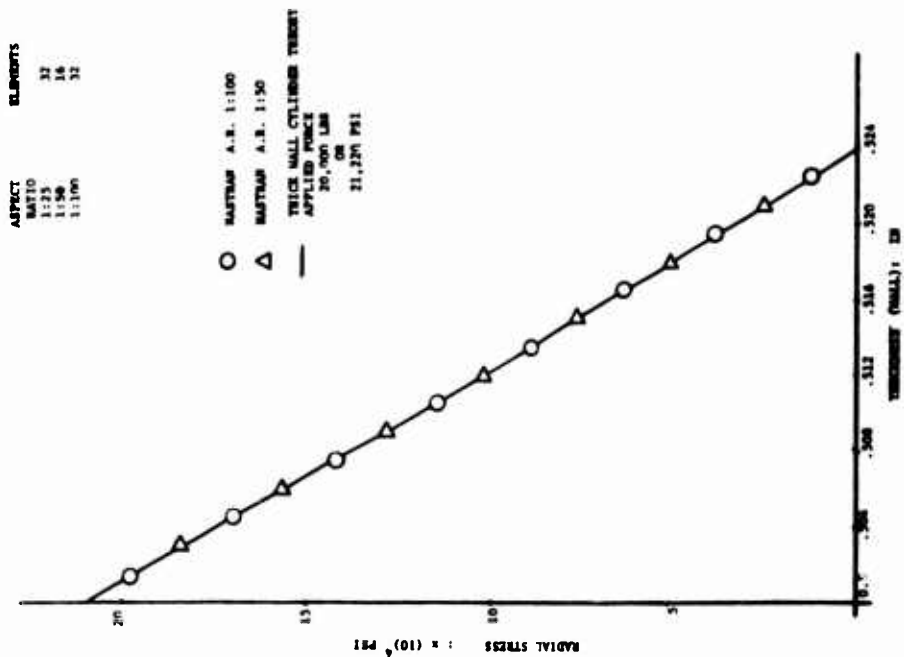


NASTRAN Model, Thin Wall Cylinder, Piece-Wise Linear Analysis

Figure IV



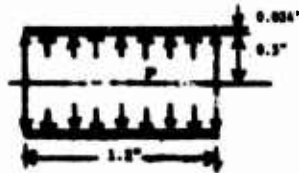
Circumferential Stress vs Wall Thickness, Thin Wall Cylinder, Aspect Ratio Study



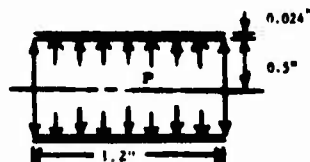
Radial Stress vs Wall Thickness, Thin Wall Cylinder, Aspect Ratio Study

Figure V

	72	63	54	45	36	27	18	9	0.524 "
81	64	56	48	40	32	27	14	8	
80	71	62	53	44	35	26	17	9	0.521 "
79	63	70	55	41	32	25	15	7	0.518 "
78	62	69	54	46	38	23	14	6	0.515 "
77	61	68	53	45	37	24	13	5	0.512 "
76	60	67	52	44	36	23	12	4	0.509 "
75	59	66	51	43	35	22	11	3	0.506 "
74	58	65	50	42	34	21	10	2	0.503 "
73	57	64	49	41	33	20	9	1	0.5 "
	1.2 "	.975 "	.75 "	.6 "	.45 "	.325 "	.225 "	0 "	



	72	63	54	45	36	27	18	9	0.524 "
81	64	56	48	40	32	24	16	8	
80	63	70	55	41	32	23	15	7	0.521 "
79	62	69	54	46	38	22	14	6	0.518 "
78	61	68	53	45	37	21	13	5	0.515 "
77	60	67	52	44	36	20	12	4	0.512 "
76	59	66	51	43	35	19	11	3	0.509 "
75	58	65	50	42	34	18	10	2	0.506 "
74	57	64	49	41	33	17	9	1	0.503 "
73									0.5 "
	1.2 "	.975 "	.825 "	.675 "	.6 "	.525 "	.375 "	.225 "	0 "



NASTRAN Models, Thin Wall Cylinder, Element Size Study

Figure VI

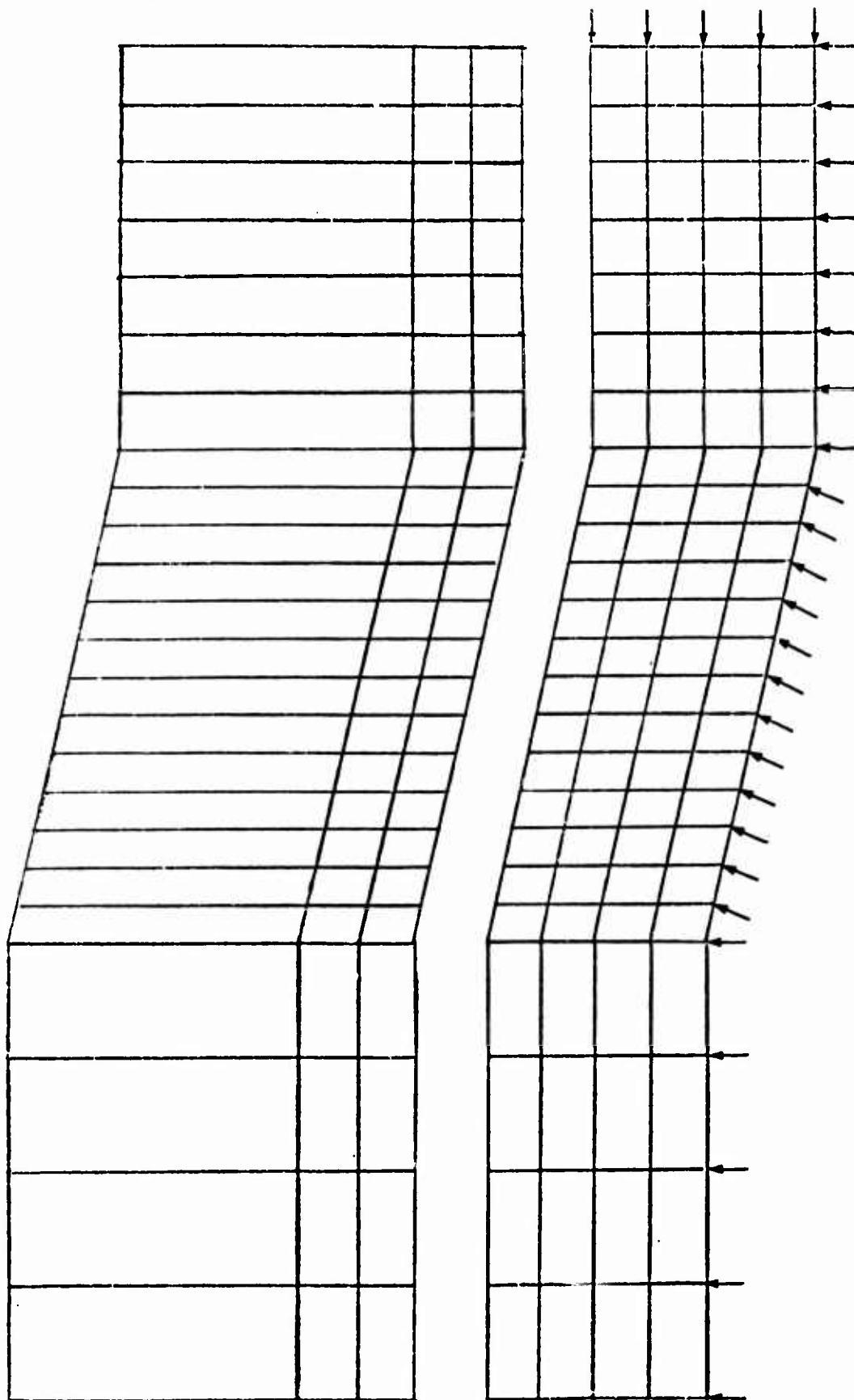


FIGURE VII
CASE NECK TAPER STUDY

ON THE ARITHMETIC MEANS AND VARIANCES OF
PRODUCTS AND RATIOS OF RANDOM VARIABLES

FRED FRISHMAN, DR.
U. S. ARMY RESEARCH OFFICE
DURHAM, NORTH CAROLINA

1. INTRODUCTION

In the sciences and other disciplines, it is quite common to encounter situations where random variables appear in combinations. Two combinations are the product and the ratio of two random variables. Given such combinations, it is required to know the mean and the variance and it is preferable that the mean and variance be determined from the original random variables. Thus, if $U = XY$ and $V = X/Y$ then we want to know the arithmetic means and variances of U and V based on our knowledge of X and Y .

2. ON THE PRODUCT OF TWO RANDOM VARIABLES

The mean and the variance of a product are well known [5]. To obtain the result in a more simple manner, we note that

$$(2.1) \quad \text{Cov}(X, Y) = E(XY) - E(X)E(Y)$$

provided that the moments exist. Henceforth, we will assume that the moments exist.

By rearranging (2.1), we obtain

$$(2.2) \quad E(XY) = \text{Cov}(X, Y) + E(X)E(Y).$$

Further, the definition of the linear correlation coefficient $\rho_{X,Y}$ is given by

$$(2.3) \quad \rho_{X,Y} = \frac{\text{Cov}(X, Y)}{[\text{Var}(X)]^{1/2} [\text{Var}(Y)]^{1/2}}$$

Preceding page blank

FRISHMAN

By introducing (2.3) into (2.2), it is seen that

$$(2.4) \quad E(XY) = E(X)E(Y) + \rho_{X,Y} [\text{Var}(X)]^{1/2} [\text{Var}(Y)]^{1/2}.$$

To obtain the variance of a product of random variables, we note that

$$(2.5) \quad \text{Var}(XY) = E(X^2 Y^2) - [E(XY)]^2,$$

and that

$$(2.6) \quad \text{Cov}(X^2, Y^2) = E(X^2 Y^2) - E(X^2)E(Y^2).$$

Then from (2.1) and (2.5),

$$\begin{aligned} \text{Cov}(X^2, Y^2) &= \text{Var}(XY) + [\text{Cov}(X, Y) + E(X)E(Y)]^2 - E(X^2)E(Y^2) \\ &= \text{Var}(XY) + [\text{Cov}(X, Y)]^2 + 2 \text{Cov}(X, Y) E(X)E(Y) \\ &\quad - \text{Var}(X) \text{Var}(Y) - \text{Var}(X) [E(Y)]^2 - \text{Var}(Y) [E(X)]^2. \end{aligned}$$

Then, by rearrangement of terms,

$$(2.7) \quad \begin{aligned} \text{Var}(XY) &= \text{Cov}(X^2, Y^2) - [\text{Cov}(X, Y)]^2 - 2 \text{Cov}(X, Y) E(X)E(Y) \\ &\quad + \text{Var}(X) \text{Var}(Y) + \text{Var}(X) [E(Y)]^2 + \text{Var}(Y) [E(X)]^2. \end{aligned}$$

If X and Y are independent or uncorrelated, (2.2) and (2.7) reduce to

$$(2.8) \quad E(XY) = E(X)E(Y),$$

and

$$(2.9) \quad \text{Var}(XY) = \text{Var}(X) \text{Var}(Y) + \text{Var}(X) [E(Y)]^2 + \text{Var}(Y) [E(X)]^2.$$

In addition, if the two variables are independent or uncorrelated, then

$$(2.10) \quad \text{Var}(XY) = \text{Var}(X) \text{Var}(Y)$$

only if the means of both random variables are zero.

3. ONE SET OF RESULTS FOR THE MEAN AND VARIANCE OF A RATIO OF RANDOM VARIABLES

We can write, using (2.3)

$$(3.1) \quad \rho_{Y, 1/X} = \frac{E\left(\frac{Y}{X}\right) - E(Y)E\left(\frac{1}{X}\right)}{[\text{Var}(Y)]^{1/2} [\text{Var}\left(\frac{1}{X}\right)]^{1/2}}.$$

Then as in (2.4),

$$(3.2) \quad E\left(\frac{Y}{X}\right) = E(Y)E\left(\frac{1}{X}\right) + \rho_{Y, 1/X} [\text{Var}(Y)]^{1/2} [\text{Var}\left(\frac{1}{X}\right)]^{1/2}$$

provided that all the terms exist. In Section 5, it is shown that

$E[\frac{1}{X^r}]$ exists provided that the density function has a zero of order at least r at the origin. For example, $E[\frac{1}{X^r}]$ does not exist for the Normal Distribution.

We obtain $\text{Var}(\frac{Y}{X})$, by replacing X with $\frac{1}{X}$ in (2.7).

$$(3.3) \text{Var}(\frac{Y}{X}) = \text{Cov}(Y^2, \frac{1}{X^2}) - [\text{Cov}(Y, \frac{1}{X})]^2 - 2\text{Cov}(Y, \frac{1}{X})E(Y)E(\frac{1}{X}) \\ + \text{Var}(Y)\text{Var}(\frac{1}{X}) + \text{Var}(Y)[E(\frac{1}{X})]^2 + \text{Var}(\frac{1}{X})[E(Y)]^2.$$

If Y and $\frac{1}{X}$ are uncorrelated, or if Y and X are independent (see Section 5), then

$$(3.4) E(\frac{Y}{X}) = E(Y)E(\frac{1}{X})$$

and

$$(3.5) \text{Var}(\frac{Y}{X}) = \text{Var}(Y)\text{Var}(\frac{1}{X}) + \text{Var}(Y)E(\frac{1}{X})^2 + \text{Var}(\frac{1}{X})[E(Y)]^2,$$

under the usual assumptions of existence.

If Y and X are uncorrelated Normals, $E(\frac{Y}{X})$ and $\text{Var}(\frac{Y}{X})$ do not exist. The proof of the lack of existence of the mean is shown in Section 5. Since the mean does not exist, neither does the variance.

In all of the above, we note the requirement for information about the reciprocal of the denominator term.

4. ANOTHER RESULT FOR THE MEAN AND VARIANCE OF A RATIO OF RANDOM VARIABLES.

By using (2.3) we can write

$$(4.1) \rho_{X, Y/X} = \frac{E(Y) - E(X)E(\frac{Y}{X})}{[\text{Var}(X)]^{1/2}[\text{Var}(\frac{Y}{X})]^{1/2}}$$

Then, by rearrangement of terms,

$$(4.2) E(\frac{Y}{X}) = \frac{E(Y)}{E(X)} - \frac{\rho_{X, Y/X}[\text{Var}(X)]^{1/2}[\text{Var}(\frac{Y}{X})]^{1/2}}{E(X)} \\ = \frac{E(Y)}{E(X)} - \frac{\text{Cov}(X, \frac{Y}{X})}{E(X)},$$

provided that the terms exist.

(Also from (2.7), where we replace Y by $\frac{Y}{X}$,

$$(4.3) \text{Var}(Y) = \text{Cov} \left[X^2, \left(\frac{Y}{X} \right)^2 \right] - \left[\text{Cov} \left(X, \frac{Y}{X} \right) \right]^2 -$$

$$2 \text{Cov} \left(X, \frac{Y}{X} \right) E(X) E \left(\frac{Y}{X} \right) + \text{Var}(X) \text{Var} \left(\frac{Y}{X} \right) +$$

$$\text{Var}(X) \left[E \left(\frac{Y}{X} \right) \right]^2 + \text{Var} \left(\frac{Y}{X} \right) [E(X)]^2.$$

Rearrangement gives

$$(4.4) \text{Var} \left(\frac{Y}{X} \right) = \frac{\text{Var}(Y) - \text{Cov} \left[X^2, \left(\frac{Y}{X} \right)^2 \right] + \left[\text{Cov} \left(X, \frac{Y}{X} \right) \right]^2}{\text{Var}(X) + [E(X)]^2}$$

$$+ \frac{2 \text{Cov} \left(X, \frac{Y}{X} \right) E(X) E \left(\frac{Y}{X} \right) - \text{Var}(X) \left[E \left(\frac{Y}{X} \right) \right]^2}{\text{Var}(X) + [E(X)]^2}.$$

Introducing (4.2) into (4.4), we get

$$(4.5) \text{Var} \left(\frac{Y}{X} \right) = \left\{ \text{Var}(Y) - \text{Cov} \left(X^2, \frac{Y^2}{X^2} \right) - \left[\text{Cov} \left(X, \frac{Y}{X} \right) \right]^2 + \right.$$

$$\left. 2 E(Y) \text{Cov} \left(X, \frac{Y}{X} \right) - \frac{\text{Var}(X)}{[E(X)]^2} \left[E(Y) - \text{Cov} \left(X, \frac{Y}{X} \right) \right]^2 \right\} \frac{1}{\text{Var}(X) + [E(X)]^2}$$

If X and $\frac{Y}{X}$ are uncorrelated,

$$(4.6) E \left(\frac{Y}{X} \right) = \frac{E(Y)}{E(X)} \text{ provided that } E(X) \neq 0,$$

and

$$(4.7) \text{Var} \left(\frac{Y}{X} \right) = \frac{\text{Var}(Y) [E(X)]^2 - \text{Var}(X) [E(Y)]^2}{[E(X)]^2 \{ \text{Var}(X) + [E(X)]^2 \}},$$

provided that $E(X) \neq 0$.

Analogous with Section 3, we are highly dependent on information about relationships between the numerator random variable and the ratio of the random variables.

5. DETERMINATION OF THE MEAN OF A RATIO BY USE OF CONDITIONAL PROBABILITIES

It follows from the definition of conditional probability that

$$(5.1) E \left[\frac{Y}{X} \right] = E_Y \left\{ E_{X|Y} \left(\frac{Y}{X} | Y \right) \right\}$$

$$= E_Y \left\{ [Y] E_{X|Y} \left(\frac{1}{X} \mid Y \right) \right\} .$$

If we consider, as a special case the Bivariate Normal, we know that the marginal distributions including the conditional are Normal distributions. And for the Normal, it is well known that

$E[\frac{1}{X}]$ does not exist so that $E[\frac{Y}{X}]$ does not exist.

More generally, if X and Y are independent, then

$$E[\frac{Y}{X}] = E[Y] E[\frac{1}{X}]$$

since if Y and X are independent, then Borel functions of Y and X are independent.

The phenomenon of the occurrence of the non-existence of $E[\frac{Y}{X}]$ is portrayed more graphically in the case of independence in a paper by H. A. David [2, pp. 122-123] as follows:

Let $G(x)$ be a cumulative distribution function. Define a central moment of negative order $-r$ by

$$(5.2) \quad \mu'_{-r}(x) = \int_{-\infty}^{+\infty} x^{-r} dG(x) dx.$$

This moment exists in the discrete case and also in the continuous case provided that the density function $g(x)$ has a zero of order at least r at the origin (i.e., $\lim_{x \rightarrow 0} x^{-r} g(x) = 0$).

Consider the case where the random variables are continuous. Let Y be independent of X with density function $f(y)$; and if

$\mu'_{-r}(x)$ and $\mu'_r(y)$ exist,

$$(5.3) \quad \mu'_{-r}(x) \mu'_r(y) = \int_{-\infty}^{+\infty} x^{-r} g(x) dx \int_{-\infty}^{+\infty} y^r f(y) dy.$$

or,

$$(5.4) \quad E \left[\left(\frac{Y}{X} \right)^r \right] = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \left(\frac{Y}{X} \right)^r g(x) f(y) dx dy = E[Y^r] E \left[\frac{1}{X^r} \right].$$

In order that the density function $g(x)$ have a zero of order of least r at the origin, we are saying, in effect, that the random variable is either all positive or all negative, and if 0 is in the range of the random variable, then $g(x)$ must be an infinitesimal at $X=0$.

6. RAO'S PROCEDURE FOR DETERMINING $E(\frac{Y}{X})$

Rao [7] determines the Expected Value of the ratio of Normally distributed variates in essentially the following fashion:

Let X and Y be bivariate normal $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$;
 $-\infty < x, y < +\infty$.

Now, let $U = \frac{Y - \mu_y}{\sigma_y}$ and $Z = \frac{X - \mu_x}{\sigma_x}$, so that $Y = \mu_y + U \sigma_y$
 and $X = \mu_x + Z \sigma_x$. From the definitions of U and Z , we have,
 $E(U) = E(Z) = 0$ and $\sigma_u^2 = \sigma_z^2 = 1$.

Then,

$$(6.1) \quad \frac{Y}{X} = \frac{\mu_y + U \sigma_y}{\mu_x + Z \sigma_x} = \mu_y \left(\frac{\mu_y + U \sigma_y}{\mu_y} \right) \frac{1}{\mu_x \left(\frac{\mu_x + Z \sigma_x}{\mu_x} \right)}$$

$$= \frac{\mu_y}{\mu_x} \left(1 + \frac{U \sigma_y}{\mu_y} \right) \left(1 + \frac{Z \sigma_x}{\mu_x} \right)^{-1}.$$

Rao expanded the last term under the assumption of convergence of the sum. It is well known, however, that $\frac{1}{1+a}$ converges provided that $-1 < a < +1$. This requirement in our case is that $-1 < \frac{Z \sigma_x}{\mu_x} < 1$ or equivalently $|Z \sigma_x| < |\mu_x|$. In terms of the original random variable, the requirement is that $-1 < \frac{X - \mu_x}{\mu_x} < +1$ or that $0 < X < 2\mu_x$ for $\mu_x > 0$, and that $2\mu_x < X < 0$ for $\mu_x < 0$. But since $-\infty < X < +\infty$ for the Normal Distribution, the requirement is not satisfied. The procedure requires that $\mu_x, \mu_y \neq 0$, also.

Nevertheless, the procedure is appropriate for random variables which satisfy the convergence conditions so we continue the development of Rao's procedure, where Z and U are appropriate standardized random variables.

We note that finite limits of the denominator random variable are a prerequisite. As examples, the procedure is valid for appropriately parameterized uniform or binomial random variables. It probably is appropriate for cases where we say that the Normal

distribution satisfies the data, since we probably mean that a truncated Normal satisfies the data, it being rare for an item or test measurement to take on both positive and negative values. By slightly rewriting (6.1) and performing the expansion, we obtain

$$\begin{aligned}
 (6.2) \quad \frac{Y}{X} &= \frac{1}{\mu_x} \left(\mu_y + U \sigma_y \right) \left[1 - \frac{Z \sigma_x}{\mu_x} + \frac{Z^2 \sigma_x^2}{2 \mu_x} - \frac{Z^3 \sigma_x^3}{3 \mu_x} + \dots \right] \\
 &= \frac{\mu_y}{\mu_x} \left[1 - \frac{Z \sigma_x}{\mu_x} + \frac{Z^2 \sigma_x^2}{2 \mu_x} - \frac{Z^3 \sigma_x^3}{3 \mu_x} + \dots \right] \\
 &\quad + \frac{\sigma_y}{\mu_x} \left[U - \frac{Z U \sigma_x}{\mu_x} + \frac{Z^2 U \sigma_x^2}{2 \mu_x} \dots \right]
 \end{aligned}$$

Taking Expected Values, we obtain

$$\begin{aligned}
 (6.3) \quad E\left[\frac{Y}{X}\right] &= \frac{\mu_y}{\mu_x} \left\{ 1 + \frac{\sigma_x^2 E(Z^2)}{2 \mu_x} - \frac{\sigma_x^3 E(Z^3)}{3 \mu_x} + \dots \right. \\
 &\quad \left. + \frac{(-1)^m \sigma_x^m E(Z^m)}{\mu^m} + \dots \right\} - \frac{\sigma_y}{\mu_x} \left\{ \frac{\sigma_x E(ZU)}{\mu_x} - \right. \\
 &\quad \left. \frac{\sigma_x^2 E(Z^2 U)}{2 \mu_x} + \dots + \frac{(-1)^{n-1} \sigma_x^n E(Z^n U)}{\mu_x^n} + \dots \right\} \\
 &= \frac{\mu_y}{\mu_x} \left\{ 1 + \sum_{i=2}^{\infty} (-1)^i \frac{\sigma_x^i E(Z^i)}{\mu_x^i} \right\} \\
 &\quad + \sigma_y \sum_{j=1}^{\infty} \frac{(-1)^j \sigma_x^j E(Z^j U)}{\mu_x^{j+1}}
 \end{aligned}$$

If, in addition to the prior conditions on the range of X and that $\mu_x \neq 0$, we specify that Z, U are bivariate central symmetrically distributed. Then from [8, pp.23], we have that

FRISHMAN

$$E(Z^{2i+1}) = 0, \quad i = 0, 1, \dots$$

and

$$E(Z^i U) = 0 \quad \text{for } i \text{ an even integer.}$$

Consequently,

$$(6.4) \quad E\left[\frac{Y}{X}\right] = \frac{\mu_Y}{\mu_X} \left\{ 1 + \sum_{j=1}^{\infty} \frac{\sigma_X^{2j} E(Z^{2j})}{\mu_X^{2j}} \right\} - \sigma_Y \left\{ \sum_{j=1}^{\infty} \frac{\sigma_X^{2j-1} E(Z^{2j-1} U)}{\mu_X^{2j}} \right\}.$$

If, in addition, we add the condition of independence between Z and U or X and Y , then (6.4) reduces to

$$(6.5) \quad E\left[\frac{Y}{X}\right] = \frac{\mu_Y}{\mu_X} \left\{ 1 + \sum_{i=1}^{\infty} \frac{\sigma_X^{2i} E(Z^{2i})}{\mu_X^{2i}} \right\},$$

which exists under the usual conditions.

We note that (6.3), (6.4) and (6.5) presented above could have been obtained by use of a Taylor Expansion of $\frac{Y}{X}$, the classical Propagation of Errors procedure.

We conclude this Section by noting that Rao's procedure for determining $E\left[\frac{Y}{X}\right]$ is valid only under very special circumstance. These are not satisfied by the Bivariate Normal.

7. THE VARIANCE OF $\frac{Y}{X}$, FOLLOWING FROM RAO'S PROCEDURE

Following the classical definition, we can write

$$(7.1) \quad \text{Var}\left(\frac{Y}{X}\right) = E\left[\frac{Y}{X} - E\left(\frac{Y}{X}\right)\right]^2.$$

For the terms in the square brackets, we utilize (6.2) and (6.3) and after extensive algebraic simplification, we arrive at

$$(7.2) \quad \text{Var}\left(\frac{Y}{X}\right) = \frac{\mu_Y^2}{\mu_X^2} \left\{ \frac{\sigma_X^2}{\mu_X^2} + \sum_{i=2}^{\infty} \frac{\sigma_X^{2i}}{\mu_X^{2i}} \text{Var}(Z^i) + \right.$$

$$\begin{aligned}
& 2 \sum_{i=k}^{\infty} \sum_{\substack{k=2 \\ i < k}}^{\infty} \frac{\sigma_x^{i+k}}{\mu_x^{i+k}} (-1)^{i+k} \text{Cov}(Z^i, Z^k) \Big\} \\
& + \frac{2 \mu_y \sigma_y}{\mu_x^2} \left\{ (-1) \frac{\sigma_x}{\mu_x} \text{Cov}(Z, U) + \right. \\
& \left. \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} (-1)^{i+j} \frac{\sigma_x^{i+j}}{\mu_x^{i+j}} \text{Cov}(Z^i, Z^j U) \right\} + \\
& \frac{\sigma_y^2}{\mu_x^2} \left\{ 1 + \sum_{j=1}^{\infty} \frac{\sigma_x^{2j}}{\mu_x^{2j}} \text{Var}(Z^j U) \right. \\
& \left. + 2 \sum_{j=0}^{\infty} \sum_{\substack{m=1 \\ j < m}}^{\infty} \frac{\sigma_x^{j+m}}{\mu_x^{j+m}} (-1)^{j+m} \text{Cov}(Z^j U, Z^m U) \right\}
\end{aligned}$$

where, for example, $\text{Var}(Z^j U) = E[Z^j U - E(Z^j U)]^2$

$$\text{Cov}(Z^j U, Z^m U) = E[Z^j U - E(Z^j U)][Z^m U - E(Z^m U)]$$

$$\text{Cov}(Z, U) = E[Z - E(Z)][U - E(U)] = E(ZU).$$

If we take only the first term in each of the curly brackets, we obtain an approximation

$$\begin{aligned}
(7.3) \quad \text{Var}\left(\frac{Y}{X}\right) & \approx \frac{\mu_y^2 \sigma_x^2}{\mu_x^4} - 2 \frac{\rho_{uz} \mu_y \sigma_y \sigma_x}{\mu_x^3} + \frac{\sigma_y^2}{\mu_x^2} \\
& = \frac{\mu_y^2 \sigma_x^2 + \mu_x^2 \sigma_y^2 - 2 \rho_{uz} \mu_y \sigma_y \sigma_x}{\mu_x^4}
\end{aligned}$$

which can be obtained, also, by the usual Propagation of Errors procedures.

If we specify that X and Y (and consequently U and Z) are independent, then (7.2) reduces to

$$(7.4) \quad \text{Var}\left(\frac{Y}{X}\right) = \frac{\mu_y^2}{\mu_x^2} \left\{ \frac{\sigma_x^2}{\mu_x^2} + \sum_{i=2}^{\infty} \frac{\sigma_x^{2i}}{\mu_x^{2i}} \text{Var}(Z^i) + \right. \\ \left. 2 \sum_{i=1}^{\infty} \sum_{\substack{k=2 \\ i < k}}^{\infty} \frac{\sigma_x^{i+k}}{\mu_x^{i+k}} (-1)^{i+k} \text{Cov}(Z^i, Z^k) \right\} \\ + \frac{\sigma_y^2}{\mu_x^2} \left\{ 1 + \sum_{j=1}^{\infty} \frac{\sigma_x^{2j}}{\mu_x^{2j}} \text{Var}(Z^j U) \right\} ,$$

and the approximation reduces to

$$(7.5) \quad \text{Var}\left(\frac{Y}{X}\right) \approx \frac{\mu_y^2 \sigma_x^2 + \mu_x^2 \sigma_y^2}{\mu_x^4}$$

8. THE DISTRIBUTION OF THE RATIO OF NORMAL VARIATES

C. C. Craig [1, pp. 24-32] presents Fieller's development [3, pp. 424-440] on the distribution of a ratio and Geary's approximation [4, pp. 442-460]. The following is a summary of the results:

Assume (X, Y) is Bivariate Normal or BVN $(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$.

Then

$$(8.1) \quad f(x, y) = \frac{1}{2\pi(1-\rho^2)^{1/2}\sigma_x\sigma_y} e^{-1/2(1-\rho^2)\left\{\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - 2\rho\left(\frac{x-\mu_x}{\sigma_x}\right)\left(\frac{y-\mu_y}{\sigma_y}\right) + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right\}} .$$

Let $v = \frac{y}{x}$ and let $x = x$. Then, $dv dx = \frac{1}{|x|} dx dy$. We define

$$(8.2) \quad g(v) = \int_{-\infty}^{+\infty} |x| f(x, xv) dx = \int_0^{\infty} x f(x, xv) dx - \int_{-\infty}^0 x f(x, xv) dx \\ = \int_{-\infty}^{+\infty} x f(x, xv) dx - 2 \int_{-\infty}^0 x f(x, xv) dx$$

Letting

$$Q(v) = \int_{-\infty}^{+\infty} x f(x, xv) dx, \text{ and } R(v) = \int_{-\infty}^0 -2x f(x, xv) dx, \text{ then}$$

FRISHMAN

$$(8.3) \quad g(v) = Q(v) + R(v)$$

where $R(v) > 0$ since $-2 \int_{-\infty}^0 x f(xv, x) dx > 0$.

First, it can be shown that

$$(8.4) \quad Q(v) = \frac{\mu_x \sigma_y^2 - \rho \sigma_x \sigma_y (\mu_y + v \mu_x) + v \mu_y \sigma_x^2}{\sqrt{2\pi} (\sigma_y^2 - 2\rho v \sigma_x \sigma_y + v^2 \sigma_x^2)^{3/2}} \\ - \frac{1}{2} \left\{ \frac{(\mu_y - v \mu_x)^2}{\sigma_y^2 - 2\rho v \sigma_x \sigma_y + v^2 \sigma_x^2} \right\}.$$

Now, $R(v)$ is of the same form as $Q(v)$ with the exception that the limits of integration are different. Thus

$$(8.5) \quad R(v) = \frac{\sigma e^{-\frac{1}{2\sigma^2} \left(\frac{C}{A} \right)}}{\pi \sqrt{A}} - 2Q(v) \Phi \left(\frac{B}{A\sigma} \right),$$

$$\text{where } \Phi \left(\frac{-B}{A\sigma} \right) = \int_{-\infty}^{-\frac{B}{A\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Now, returning to the distribution of the ratio (8.3), $g(v) = Q(v) + R(v)$, we replace $R(v)$ by the result in (8.5) to obtain

$$(8.6) \quad g(v) = Q(v) - 2Q(v) \Phi \left(\frac{-B}{A\sigma} \right) + \frac{\sigma e^{-\frac{1}{2\sigma^2} \left[\frac{C}{A} \right]}}{\pi \sqrt{A}} \\ = Q(v) \left\{ 1 - 2\Phi \left(\frac{-B}{A\sigma} \right) \right\} + \frac{\sigma e^{-\frac{1}{2\sigma^2} \left[\frac{C}{A} \right]}}{\pi \sqrt{A}}$$

Thus we have the distribution of the ratio in closed form, where $Q(v)$ is given by (8.4), and

$$(8.7) \quad \left\{ \begin{array}{l} A = \sigma_y^2 - 2\rho v \sigma_x \sigma_y + v^2 \sigma_x^2 \\ B = \mu_x \sigma_y^2 + v \mu_y \sigma_x^2 - \rho(\mu_y + v \mu_x) \sigma_x \sigma_y \\ C = \mu_x^2 \sigma_y^2 - 2\rho \sigma_x \sigma_y \mu_x \mu_y + \mu_y^2 \sigma_x^2 \\ \sigma^2 = \frac{(1 - \rho^2) \sigma_x^2 \sigma_y^2}{A} \end{array} \right.$$

we can rewrite $Q(v)$ by using (8.7) as follows:

$$(8.8) \quad Q(v) = \frac{B}{\sqrt{2\pi A}^{3/2}} e^{-\frac{1}{2\sigma^2} \left(\frac{AC - B^2}{A^2} \right)}.$$

Now, we return to (8.5), and (8.7). We note that as $\mu_x \rightarrow \infty$, then $B \rightarrow \infty$ and $C \rightarrow \infty$. As a result,

$$(8.9) \quad e^{-\frac{1}{2\sigma^2} \frac{C}{A}} \rightarrow 0,$$

and

$$\Phi\left(-\frac{B}{A\sigma}\right) \rightarrow 0.$$

From this, we conclude that

$$(8.10) \quad R(v) \rightarrow 0 \text{ as } \mu_x \rightarrow \infty.$$

Therefore, for μ_x sufficiently large, we can get a good approximation of $g(v)$ by $Q(v)$, only.

Thus, for the case where μ_x is large, from (8.4) we have

$$(8.11) \quad Q(v) = \frac{\mu_x \sigma_y^2 - \rho \sigma_x \sigma_y (\mu_y + v \mu_x) + v \mu_y \sigma_x^2}{\sqrt{2\pi} (\sigma_y^2 - 2\rho v \sigma_x \sigma_y + v^2 \sigma_x^2)^{3/2}} e^{-\frac{1}{2} \left\{ \frac{(\mu_y - v \mu_x)^2}{\sigma_y^2 - 2\rho v \sigma_x \sigma_y + v^2 \sigma_x^2} \right\}}$$

We impose the transformation

$$(8.12) \quad z = \frac{v \mu_x - \mu_y}{(\sigma_y^2 - 2\rho v \sigma_x \sigma_y + v^2 \sigma_x^2)^{1/2}}, \text{ with } dz = \frac{B}{A^{3/2}} dv,$$

FRISHMAN

and we note that the denominator of $Q(v)$ is positive for all v .

Hence

$$(8.13) \quad Q(z) = Q \left[\frac{v\mu_x - \mu_y}{(\sigma_y^2 - 2\rho v\sigma_x\sigma_y + v^2\sigma_x^2)^{1/2}} \right].$$

It follows that, as $\mu_x \rightarrow \infty$ for σ finite, z approaches a $N(0,1)$. This is the Geary result [4].

If in (8.6), $\mu_x = \mu_y = 0$, we obtain

$$(8.14) \quad g_1(v) = \frac{(1 - \rho^2)^{1/2}}{\pi \left(\frac{\sigma_y}{\sigma_x} - 2\rho v + v^2 \frac{\sigma_x}{\sigma_y} \right)},$$

a generalized Cauchy distribution.

The cumulative distribution function for (8.14) is

$$(8.15) \quad G(v) = \int_{-\infty}^v g(v) dv = \frac{1}{\pi} \arctan \left[\frac{\sigma_x v - \rho \sigma_y}{\sigma_y (1 - \rho^2)^{1/2}} \right]$$

as obtained from Gradshteyn and Ryzhik, [6, p. 82].

The generalized Cauchy obtained by this writer is a simple extension and has utility more from a theoretical viewpoint than an applications viewpoint.

However, the Geary result has great applicability as do the approximation results in Sections 6 and 7 when one is willing to assume that X and Y are Bivariate Normal. As noted in Section 6, in the real world, few if any, random variables are truly normally distributed since few, if any, can range from $-\infty$ to $+\infty$. Few, if any, will take on both positive and negative values! Thus the mathematical abstraction, the Normal Distribution, can be at best a good approximation, and a good approximation only around the mean of the data, since ordinarily one does not see data points in the extreme tails of the Distribution.

9. SUMMARY

In Sections 3 and 4, exact expressions are given for the Mean and the Variance of a Ratio of random variables, regardless of the form of the random variables. One problem is that certain correlation coefficients must be known or assumed to be zero in order to use the formulae. Also, there is the usual requirement that the moments exist. Based on the results of Section 5, we know that these exact results are not appropriate for the Normal Distribution and those other distributions where the random variable takes on negative and positive values and/or the value 0 with probability greater than 0.

In Section 6, a procedure is given for getting the Mean of a Ratio. The requirement is that the denominator random variable be such that either $2\mu_x < X < 0$ when μ_x is negative or that $0 < X < 2\mu_x$ when μ_x is positive. Also, we require that the Mean of the numerator $\mu_y \neq 0$. In Section 7, we obtain the Variance. Since the results in both Sections require summing infinite series, approximations can be obtained by terminating the summation wherever desired.

Finally in Section 8, we sketch out the development for an approximation of the distribution of the ratio if both random variables are Normally Distributed, in the sense that the Denominator variable does not take on both positive and negative values. It is claimed that it is rare for actual random variables to be Normally distributed in the mathematical sense. Rather, the Normal Distribution is a reasonable approximation at best, and under these circumstances, the results presented in Sections 6, 7 and 8 are appropriate to be used.

10. BIBLIOGRAPHY

1. Craig, C. C. (1942) "On Frequency Distributions of the Quotient and of the Product of Two Statistical Variables", Amer. Math. Monthly, Vol. 49, No. 1.
2. David, H. A. (1955) "Moments of Negative Order and Ratio-Statistics", Journal of the Royal Stat. Society, Part 1.
3. Fieller, E. C. (1932) "The Distribution of the Index in a Normal Bivariate Population", Biometrika, Vol. 24.
4. Geary, R. C. (1932) "The Frequency Distribution of the Quotient of two Normal Variables", Journal of the Royal Statistical Society, Vol. 93.

FRISHMAN

5. Goodman, L. (1960) "On the Exact Variance of Products",
Journal of the American Stat. Assn., Vol. 55.
6. Gradstein, I. S. and Ryzhik, I. M. (1965) Tables of Integrals, Series and Products. Academic Press.
7. Rao, C. R. (1952) Advanced Statistical Methods in Biometrics Research. John Wiley and Sons, Inc.
8. Frishman, F. (1971) "On Ratios of Random Variables and an Application to a Non-linear Regression Problem". Ph.D. dissertation, The George Washington University.

TNT BY-PRODUCT ISOMER RECOVERY FOR CREDIT

EVERETT E. GILBERT, DR.
PICATINNY ARSENAL
DOVER, NJ

INTRODUCTION

For every 100 tons of specification TNT, we produce approximately 4 tons of troublesome by-product TNT isomers, which must be removed by purification. This not only results in yield loss, but contributes substantially to the bad reputation of TNT manufacture as a source of air and water pollution. The isomers are removed from the crude TNT by treatment with aqueous sodium sulfite ("sellite"). They are thus converted to water-soluble sulfonates I and II (Figure 1) as a solution known as "red water." The relative proportions of each are indicated. At present, we make about 12 million lbs./year of these isomers.

Ever since World War I when large-scale TNT manufacture began, chemists concerned with the process both in government and industry here and abroad have given thought to possible methods for converting sulfonates I and II to useful derivatives. For success in such an effort, several requirements must be met:

- (1) Chemical feasibility.
- (2) Reasonable yields and ease of operation.
- (3) Prior existence of a market for the product much larger than the amount to be produced.

PAST ATTEMPTS TO UTILIZE ISOMERS

We are aware of three past attempts to find uses for sulfonates I and II. Many aromatic sulfonic acids are converted back to the parent aromatic compound by hydrolysis. This is a general reaction which has been widely used for research and industrial purposes (1).

Preceding page blank

GILBERT

Application of it to I and II, as shown in Figure 1, would yield only one compound, 2,4-dinitrotoluene. This would of course be recycled to the manufacture of TNT, thereby increasing the yield and at the same time alleviating the disposal problem. Unfortunately, nitroaromatic sulfonic acids are apparently the only type which have so far eluded attempts at this type of hydrolysis, thus disappointing several successive generations of TNT chemists who have independently had this "bright idea" for an ideal solution to the problem.

The second and third attempts involved conversion of the isomers to methylpicric acid and methyltetryl, respectively, in a manner which would satisfy criteria (1) and (2) above. Unfortunately, neither of these materials satisfied the third requirement, and so these processes were not taken seriously.

At present the red water is given without charge to paper companies who "furnace" it for the recovery of crude sodium sulfate. This practice is objectionable in that the organic values are lost during furnacing, and their decomposition results in considerable atmospheric pollution. Furthermore, there is some doubt as to whether the paper companies will continue to take this material in the future as they have in the past. The government must therefore be in a position to dispose of the red water if such should occur.

DISCUSSION OF NEW PROCESS

We report herein a new approach to the recovery of I and II in a form which could reflect substantial monetary credit to the process. The chemical steps involved, which we have demonstrated experimentally in the laboratory, are summarized schematically in Figure 2. The two steps are (1) reduction of the nitro groups of I and II to amino groups and (2) removal of the sulfonic groups in III and IV by hydrolysis which yields one product V. We will now discuss the procedure in more detail.

The red water would first be acidified with sulfuric acid. This converts residual sodium sulfite to sodium sulfate with liberation of SO_2 , which would be recovered and recycled. Non-sulfur-containing organic, including alpha-TNT, trinitrobenzoic acid, trinitrobenzaldehyde, trinitrobenzyl alcohol, and trinitrobenzene, would also be liberated and can be recovered for credit by filtration and/or solvent extraction of the acid solution. (The last four materials are of possible interest for a new procedure for making trinitrobenzene (2)). The aqueous solution now contains I, II, sodium sulfate, and impurities.

The next step involves the reduction of I to III and II to IV as shown in Figure 2. The reduction of nitro to amino groups is,

GILBERT

of course, a standard industrial operation and can be done in various ways. The conversion of pure II to IV, using iron and hydrochloric acid, is known (3). The most modern approach involves catalytic hydrogenation, which is the method used today for converting mixed 2,4-dinitrotoluenes to the corresponding diamines on a large commercial scale. Catalytic reduction of nitroaromatic sulfonic acids was first reported in 1951 (4), and the work was limited to the mononitro compounds. We therefore felt it desirable to demonstrate this in the case of a dinitroaromatic sulfonic acid analogous to I and II. Accordingly, we have shown that pure aqueous sodium dinitrobenzenesulfonate hydrogenates easily to the corresponding diamino compound using Raney nickel catalyst. Since this compound is the same as I except for the presence of the methyl group, we anticipate no difficulty in obtaining good yields in this step.

Using purchased III, which is used commercially as a dye intermediate, we have found that as the sodium salt it is very soluble in water. However, upon acidification the acid form of III precipitates quantitatively and can be recovered easily by filtration.

The next step involves the hydrolysis of III and IV to V. It is noted that this type of reaction is inoperable with I and II, as discussed above (cf. Figure 1). However, the literature does report the hydrolysis of two monoaminoaromatic sulfonic acids (5), although no product was obtained from a third and no work has apparently been done on the hydrolysis of diaminoaromatic sulfonic acids such as III or IV. We have, therefore, studied the hydrolysis of purchased III (after conversion to its acid form) to V. We have found that hydrolysis does occur upon refluxing (150°C) for several hours with 65% sulfuric acid. This procedure yields V as the sulfate salt, which crystallized from the acid upon cooling and was recovered by filtration. The acid filtrate is made up to the original weight with fresh 65% acid, and the hydrolysis cycle is repeated. We have tested this system in 5 cycles, and the final filtrate acid appears suitable for continued recycle. The solid amine salt is dissolved in water and neutralized for the separation of the free base V, which is filtered as a solid. The yield of V is 90%.

Compound III comprises two-thirds of the diaminosulfonic acid mixture, the other third being IV. Unlike III, IV is not commercially available and is accessible only by synthesis. We have, therefore, not studied the hydrolysis of IV, but we predict that it should behave like III in undergoing conversion to V.

GILBERT

PRODUCT DISPOSAL

The third requirement of any by-product recovery process, as discussed earlier, stipulates that the final product V be readily disposable. Our laboratory data indicate the formation of a possible several million lbs/year of V. This compound is widely and increasingly used to make polyurethane resins. In 1972 the production was spread over 7 companies and was reported to be ca. 500 million lbs/year with a 15% annual increase (6). Thus, the total amount of V obtainable from all of the current government TNT production would comprise only a small percent of the amount produced industrially, and there should, therefore, be no difficulty in disposing of such a relatively small quantity.

Disposal of V should result in monetary credit to the process. The diaminotoluene produced commercially for polyurethane resins usually comprises an approximately 75-25 wt % mixture of the 2,4- and 2,6-isomers, based on the product obtained by dinitrating toluene. The pure 2,4-isomer is preferred, but of course more expensive than the mixture. The V obtained by the present approach would, therefore, have an advantage over the commercial mixture, since our material would be 95% pure or better. Commercial production cost of this diaminotoluene mixture is about 18-19¢/lb, with cost of the pure 2,4-isomer (V) being somewhat higher (7). If we assume a selling price of 15¢/lb for it, a substantial credit results based on the production of several million lbs/year of V (possibly as much as \$900,000 per year).

APPLICATION TO NEW TNT PROCESS

The above line of thought applies to the existing TNT process. We are currently developing an improved process developed by Stanford Research Institute. A major objective of the SRI study was to eliminate entirely the formation of meta isomers. Success in this effort would eliminate red water and the need for expedients such as that described herein. It was found that by using refrigeration in the mono-dinitration stage, the meta isomers can be reduced in proportion to the degree of cooling. To eliminate them a temperature of -110°C is required, which is impractical. Laboratory data show that at a more practical temperature, meta isomer formation can be about halved; but at the present time, we see no prospect of the complete elimination of meta isomer formation.

A second finding of the SRI study was that the process can advantageously be interrupted after the dinitration stage, if desired, and that the meta isomers can be removed from the crude DNT rather than from the crude TNT, as is now done. Would the reduction-hydrolysis recovery concept be applicable in this situation?

GILBERT

It is known (8) that two compounds, VI and VII, shown in Figure 3 comprise most of the meta isomer content of crude DNT. The literature also states (9) that "the reaction of sulphitation of DNT may be used in practice for removing the m-toluene derivatives from the mixture of dinitro isomers." Recent studies at Edgewood Arsenal (10) substantiate this statement. This reaction is involved in the conversion of VI and VII to VIII and IX.

We believe that the reduction of VIII and IX to the corresponding aminoaromatic sulfonic acids should present no major problem for the reasons mentioned above with relation to the reduction of I and II.

The hydrolysis of these sulfonic acids to X and XI is not reported in the literature, even though (as mentioned above) this type of reaction is known for two of the aminobenzenesulfonic acids. Accordingly, we made a study of the preparation of X from the amino analogue of VIII. Yields varied from about 50% (using sulfuric acid) to 85% (using trifluoromethanesulfonic acid). We conclude that the desired reaction does occur in fair yields, but that it is more sensitive to reaction conditions than the conversion of III to V.

What about disposal of the toluidines? There is much less information on this than on 2,4-toluenediamine. It is known that they are articles of commerce with several producers and with varied uses. Volume of production is unavailable but is certainly much less than that of 2,4-toluenediamine.

CONCLUSION

In conclusion, we have demonstrated on a preliminary basis the chemical feasibility of reduction-hydrolysis concept for recovery of TNT and DNT by-product meta isomers as the corresponding amines. Further study aimed toward ultimate large-scale production appears warranted.

EXPERIMENTAL SECTION

Catalytic Hydrogenation of Sodium 2,4-Dinitrobenzenesulfonate.

Sodium 2,4-dinitrobenzenesulfonate (2.5g. as a 5% aqueous solution) was treated with hydrogen gas at 600 PSIG for 1 hour at 125°C in the presence of 1.0g. Raney 28 catalyst. Pressure decrease was 340 PSIG corresponding to quantitative conversion of the nitro to amino groups (Reference: Research Notebook 762-247, pg. 70).

GILBERT

Hydrolysis of 2,4-Diaminotoluenesulfonic Acid (III)
to 2,4-Toluenediamine (V)

The acid (20g. of purified commercial material) was suspended in 200g. 65% sulfuric acid, and the mixture was refluxed with stirring for 7 hours (148 to 155°C). The solution was cooled, crystallized, and filtered to recover the solid V sulfate. The filtrate was reconstituted to 200g. by adding fresh 65% acid, and the cycle was repeated four times. The combined V sulfate cakes were dissolved in water and neutralized to liberate free V which was identified by melting point and infrared spectrum. The yield (including that present in the final filtrate) was 90% (Reference: 762-247-20 and 27). The final filtrate can presumably be recycled at least several more times.

Hydrolysis of 4-Aminotoluene-5-Sulfonic Acid to 4-Toluidine (X)

The acid (20g.) was refluxed 7 hours at 178-82° with 200g. 75% sulfuric acid. Neutralization gave a 56% yield of 4-toluidine identified by melting point and infrared spectrum. Phosphoric and trifluoromethanesulfonic acids gave yields of 74 and 84%, respectively (Reference: 762-247-35 and 39).

GILBERT

REFERENCES

1. E. E. Gilbert, "Sulfonation and Related Reactions," Chapter 8, Wiley-Interscience, New York, 1965.
2. E. E. Gilbert, "TNB Recovery from Spent Sellite," file memorandum dated 27 March 1973.
3. F. H. Gornall and R. Robinson, J. Chem. Soc. 1926, 1981.
4. F. Allison, J. L. Comte, and H. E. Fierz-David, Helv. Chim. Acta 34, 818 (1951).
5. A. A. Spryskov and N. A. Ovsyankina, Obshchei Khim. 24, 1810 (1954); Chem. Abstracts 49, 12342 (1955).
6. Chemical and Engineering News, 1 January 1973.
7. Stanford Research Institute, private communication.
8. A. Lonnfors and B. Jansson, Chem. Eng. World 7(2), 33 (1972).
9. T. Urbanski, "Chemistry and Technology of Explosives," Vol. 1, pg. 284, Pergamon Press, New York, 1965.
10. D. H. Rosenblatt and W. H. Dennis, Jr., "Reactions of Nucleophiles," Progress Report for April 1972 under PRON GG02598601, Edgewood Arsenal, MD.

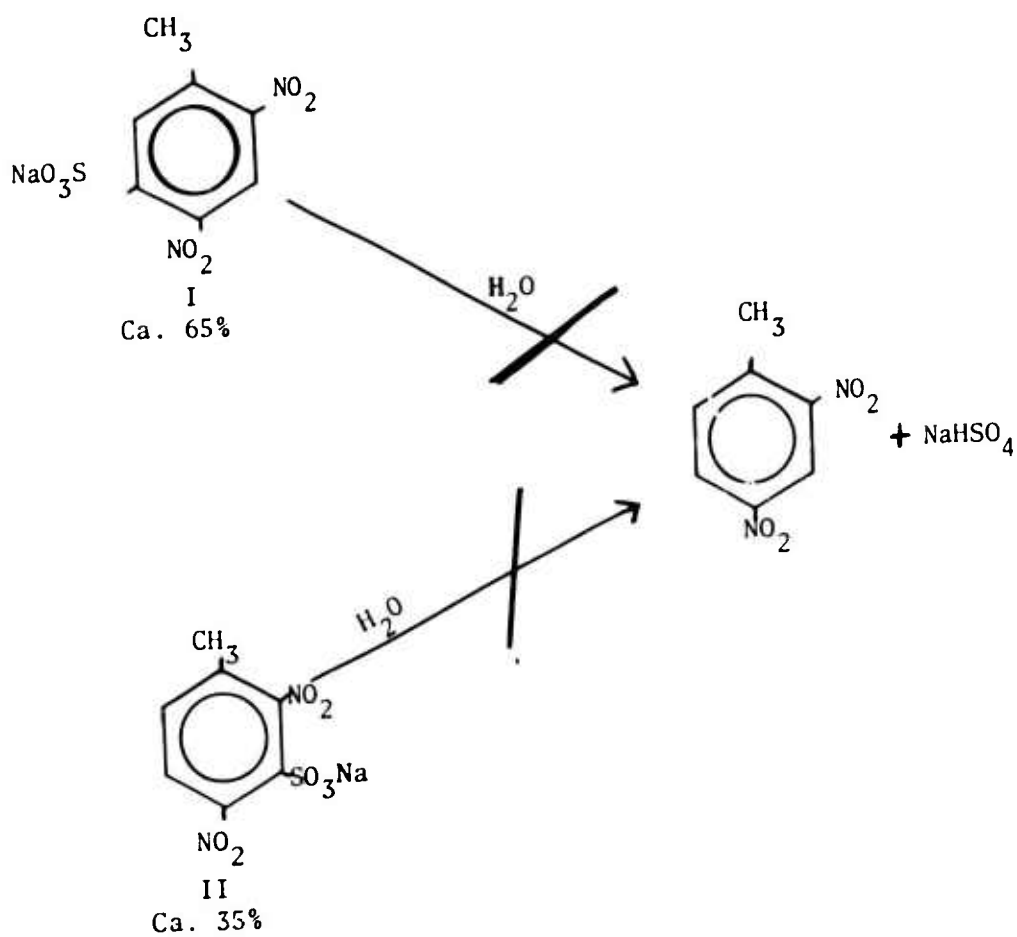


Figure 1
Scheme for Proposed Recovery of 2,4-DNT from Spent Sellite

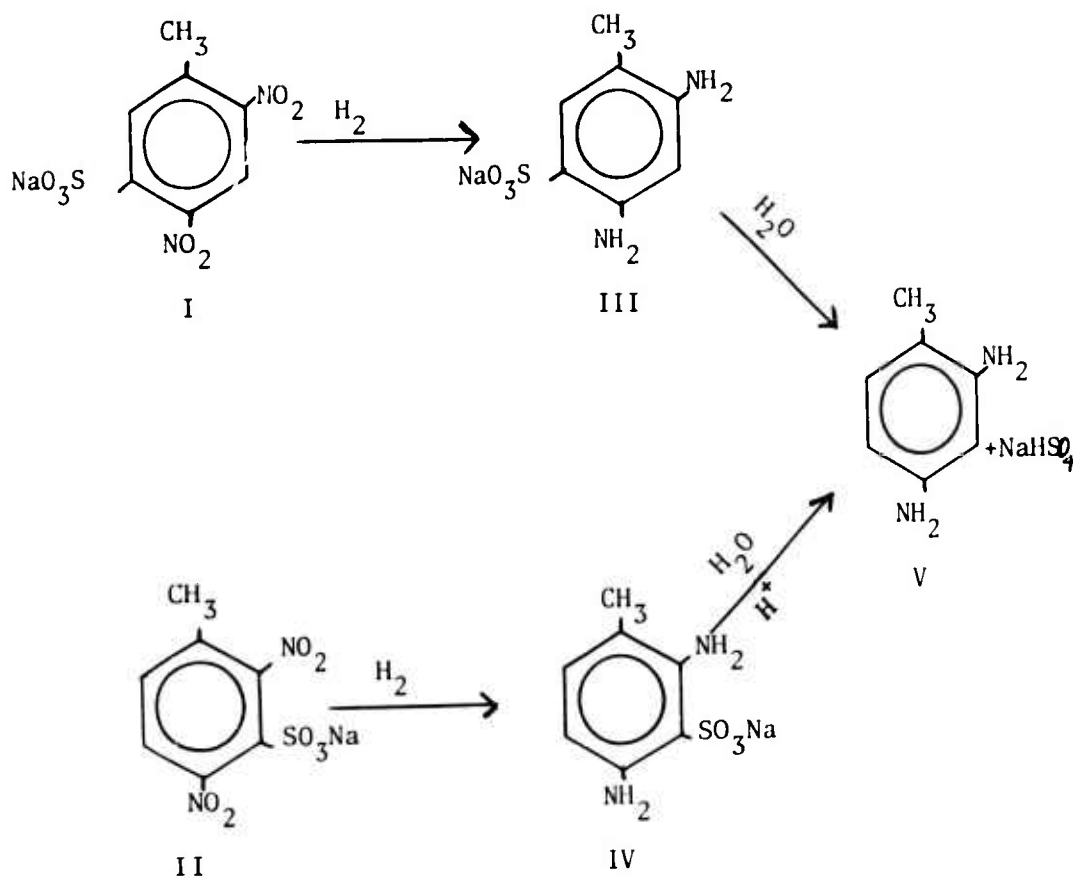


Figure 2
Scheme for Recovery of 2,4- Diaminotoluene

GILBERT

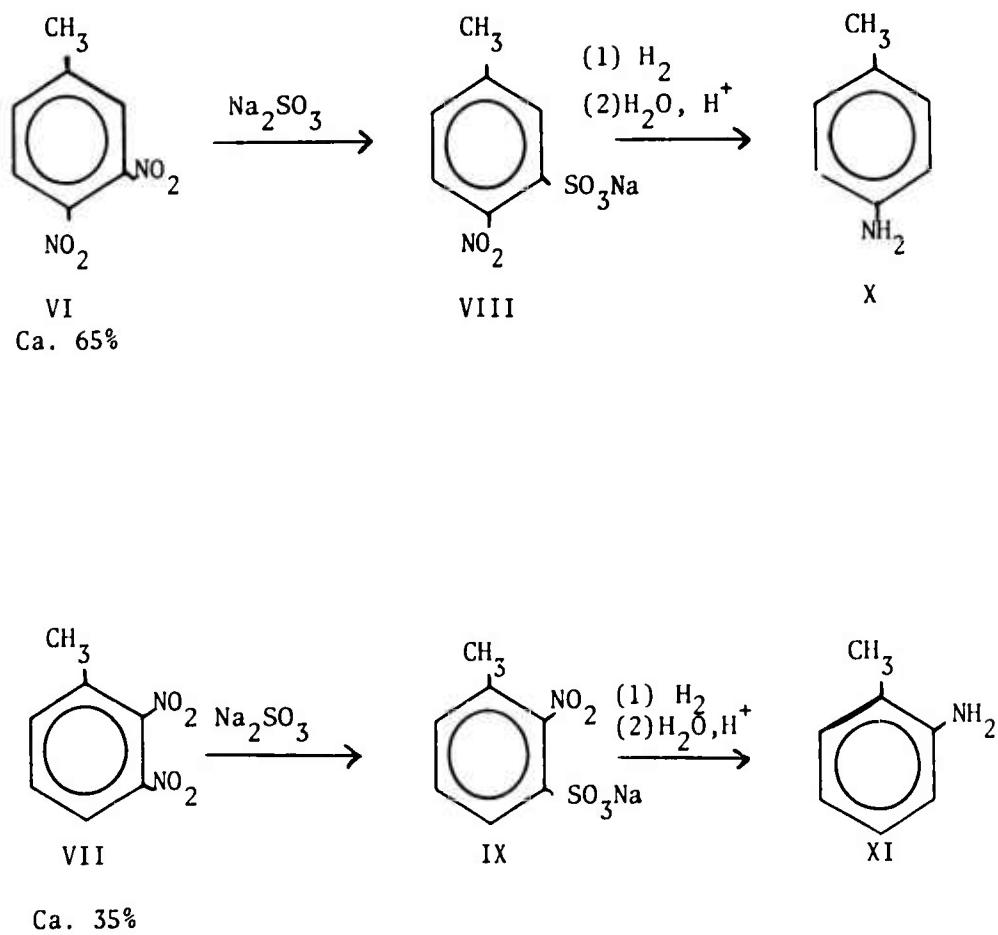


Figure 3

Scheme for Recovery of Toluidines

RECENT TRANSLOCATION RESULTS USING NAVIGATION SATELLITES

FREDERICK M. GLOECKLER, JR., MR.
U. S. ARMY ENGINEER TOPOGRAPHIC LABORATORIES
FORT BELVOIR, VIRGINIA

Translocation, as applied by users of the Navy Navigation Satellite System (NNSS), is a procedure used to position one receiver with respect to another using data acquired simultaneously by both receivers (Ref. 1). The translocation process minimizes the effect of system error sources which are highly correlated at both receiver sites.

Previously reported single-pass translocation accuracies reached a limit of about 8 meters CEP as shown in Figure 1. Figure 1 is a summary of translocation accuracy obtained in a joint Army/Marine Corps military potential test of the AN/PRN-6 (XN-1) translocation equipment (Ref. 2).

This paper presents test results for an improved class of translocation equipment. Accuracy improvement, by a factor of three, was shown for short distances.

NNSS.

The NNSS, often called TRANSIT, is an operational, world-wide, positioning system which employs earth-orbiting satellites. The satellites are in circular polar orbits at an altitude of approximately 1100 Km. The satellites transmit two stable carrier frequencies (400 and 150 MHz) modulated by a digital message containing satellite ephemeris and system timing information.

A network of ground stations tracks the satellites to determine their precise orbits. This information is extrapolated to the

GLOECKLEP

future to obtain predicted ephemerides. The ephemeris data is relayed to each satellite where it is stored in memory. The satellites continuously transmit current ephemeris data which is updated from memory every two minutes.

The basic measurement made by a receiver is the doppler frequency on the received carrier, integrated (counted) over some time interval. It can be shown that this integrated doppler count is a measure of the change in range between the satellite and observer over the integration interval (Ref. 3). An observer's position can be calculated from a set of three or more such measurements.

Error Sources.

The primary navigation error sources are: ephemeris error, uncertainty in the navigator's elevation, uncertainty in the navigator's velocity, refraction effects, and instrumentation errors.

Ephemeris errors are the result of the inability to precisely predict satellite orbits and some satellite hardware constraints. The effect of ephemeris error limits single-pass navigation accuracy to about 30 meters CEP. Ephemeris error effects are well correlated over short distances and can be minimized by translocation.

The nearly coplanar geometry of a single satellite pass limits the useful solution of the navigation equations to two dimensions. Usually, elevation is considered known, and latitude and longitude are computed. Three-dimension solutions are effective if data from two or more passes of differing geometry are used.

Uncertainty in the navigator's velocity vector during data collection will produce position errors. This error component does not affect stationary users.

An observer equipped with a two-channel receiver can use a two-frequency compensation for ionospheric refraction error. Gross corrections for tropospheric refraction can be made using meteorological data. Refraction effects are fairly well correlated over short distances and can be minimized by translocation.

The primary instrumentation error in typical NNSS receivers is imprecise determination of the doppler count integration interval. Generally, time recovery is synchronized to the satellite digital message clock which has a 51-Hz rate. Time recovery jitter can be caused by noise, local signal fading, multipath, etc., and is not well correlated at different antenna locations. Time recovery can be

GLOECKLER

improved by measuring the integration interval using a clock driven by the receiver's stable oscillator. The doppler count is corrected for the difference between the measured interval and the expected value.

Equipment.

Two types of receivers were tested to investigate the effect of improved time recovery on translocation accuracy:

1. Magnavox AN/PRR-14 Geociever
2. Modified Honeywell AN/PRN-6 (XN-1) Translocation Receiver (Modified Backpack)

The AN/PRR-14 Geociever is an operational receiver primarily used by the Defense Mapping Agency (DMA) for geodetic survey. The Geociever provides doppler measurements on both carrier frequencies and the precise time of day when the measurement was obtained. Nominal integration intervals of 30, 60, or 120 seconds are selectable.

The AN/PRN-6 (Backpack) is an advanced development man-portable receiver developed by the Applied Physics Laboratory (APL) of Johns Hopkins University for Marine Corps evaluation. A pair of Backpacks was modified at U. S. Army Engineer Topographic Laboratories (USAETL) to improve time recovery. The Modified Backpacks measure integrated doppler on the 400 MHz. carrier and measure the integration interval. The nominal integration interval is 120 seconds.

Data Handling and Reduction.

Backpack data from the field station were transmitted via telephone line to a Honeywell H316R mini-computer at Fort Belvoir. The data were reduced automatically after each pass to provide two-dimension translocations. The program, written at USAETL, was based on the program described in Reference 4.

Geociever data were collected on-site on punched paper tape. The data were processed at the DMA Topographic Center using the geodetic reduction program "DOPPLR". The Geociever data format and data reduction process permitted the following areas to be investigated: time recovery, ionospheric refraction correction, two-dimension translocation, and three-dimension translocation.

The ephemeris broadcast by the satellite was used for all results presented in this paper. The Backpack system acquired the

GLOECKLER

ephemeris in real-time. The Geociever data were processed using ephemeral data collected by the Backpack system.

The editing criteria employed fall in the areas of; elevation-angle constraints, solution residuals, and number of doppler counts used in the solution.

Since residual refraction effects are relatively large at low-elevation angles, two low-elevation angle constraints were applied: individual doppler counts acquired at elevation angles below five degrees were deleted, and passes with elevation angles at closest approach below 15 degrees were not used. Passes with elevation angles above 65 degrees were not used for two-dimension solutions since geometric dilution becomes severe. In general, the elevation angle constraints were determined from the position error distribution.

Doppler count and ephemeris errors appear as residual range errors in the least-square solution of the navigation equations. High residuals indicate a solution may be poor. For translocation, the ephemeris error contribution to the residuals should be nearly the same at both locations. Differences in residuals can be attributed to doppler-count errors. This concept of residual balance was found useful in identifying poor solutions for the Modified Backpack system.

To provide over-determination in the solution, the minimum number of doppler counts used per pass was four.

Test Area.

Several sites in the Washington, DC, area were used in the test program. Different translocation lines were used for the Geocievers and the Modified Backpacks. The site locations are shown in Figure 2.

Test Results:

Geociever tracking data were used to investigate the effects of length of doppler count, improved time recovery, and ionospheric refraction correction on two-dimension translocation accuracy. The same set of data was used for the various investigations so that results could be compared on the same basis.

Ionospheric-refraction-corrected doppler counts were processed without time-recovery correction to simulate Backpack data. The 30-second doppler counts were sequentially added to produce 60 and 120-

GLOECKLER

second counts. The results were very similar to Backpack accuracies (11 meters CEP) with little variation versus length-of-count. Use of the 30-second interval produced about 22 percent more "good" passes than using two-minute counts.

The data were reprocessed using the time-recovery information provided by the internal clock of the Geociever. The position errors ranged from 4.4 meters CEP using 120-second intervals to 2.3 meters CEP using 30-second intervals. Clearly, there is a substantial reduction in position error using better time recovery.

The above results are summarized in Figure 3. The curves are least-square fit of form $Y = aX^b$. It is interesting to note the exponent for the improved time-recovery case is nearly one-half. This implies that statistical averaging is taking place as the number of data points increases.

When the data for the 30-second interval with time-recovery correction case were reprocessed without ionospheric refraction correction the position error grew from 2.3 meters CEP to 3.1 meters.

The Modified Backpack system permitted a large amount of data to be collected automatically. Both receivers were collocated using a common antenna to determine the system noise level. The collocation accuracy was 3.1 meters CEP.

An additional collocation experiment was performed using a common oscillator. The same position accuracy was achieved, thus indicating oscillator performance is not a limiting factor. The doppler count differences between the two receivers were evaluated. The RMS error was 1.9 counts with no time-recovery correction versus 0.7 count with the correction. The error with the correction is just slightly larger than expected from a linear distribution of a ± 1 count incrementation error. It appears that increased count measurement resolution could produce a further improvement in measurement accuracy.

Three translocation lines of length 9, 57, and 89 kilometers were used with the Modified Backpacks. The respective position accuracies were 3.3, 5.0, and 8.8 meters CEP (Figure 4). The error plot and radial-error histogram for the 57 Km. line are shown in Figures 5 and 6, respectively.

One of the problems with previous translocation systems was that "bad" solutions were sometimes difficult to identify. Of the 492 passes obtained in this test series which met the elevation

angle criteria and four-count minimum, only four passes were rejected. Two had high-resolution residuals and two had residual imbalance between the solutions for the two stations.

Figure 7 demonstrates the relationship between pass geometry, as measured by elevation angle at closest approach, and longitude error. The increased error at low angles is attributed to increased refraction errors. The error at high angles is caused by geometric dilution.

The Geociever data were used to investigate three-dimension translocation. The 30 second count interval and time-recovery correction were used. The data were processed with groups of 2, 3, 4, 6, and 8 passes in the solution to examine accuracy improvement with increased number of passes.

A group of 57 passes were processed using ionospheric refraction correction. Accuracy was on the order of 5 meters RMS in each coordinate for two passes and improved by a factor of two with eight passes. This group of passes contained some poorly distributed data. A set of 27 passes with well distributed data exhibited 1 meter RMS accuracy using 8 passes (Figure 8).

The same set of 27 passes were processed without ionospheric refraction correction. The error grew by a factor of two (Figure 9).

Conclusions.

1. Time-recovery error is a significant translocation error source. The hardware necessary to reduce this error source is very inexpensive using today's technology. Figure 10 is a prediction of attainable accuracy for future systems for both single and two-channel receivers.

2. Proper editing techniques coupled with improved receivers can reduce the level of undetected "bad" solutions to a very low level.

3. With the level of equipment performance reports herein, errors in relative height between the two translocation stations probably will be the dominant error source for an operational system employing two-dimension, single-pass solutions.

4. Three-dimension translocation to an accuracy of 5 meters RMS in each direction is easily attainable using two or three passes.

GLOECKLER

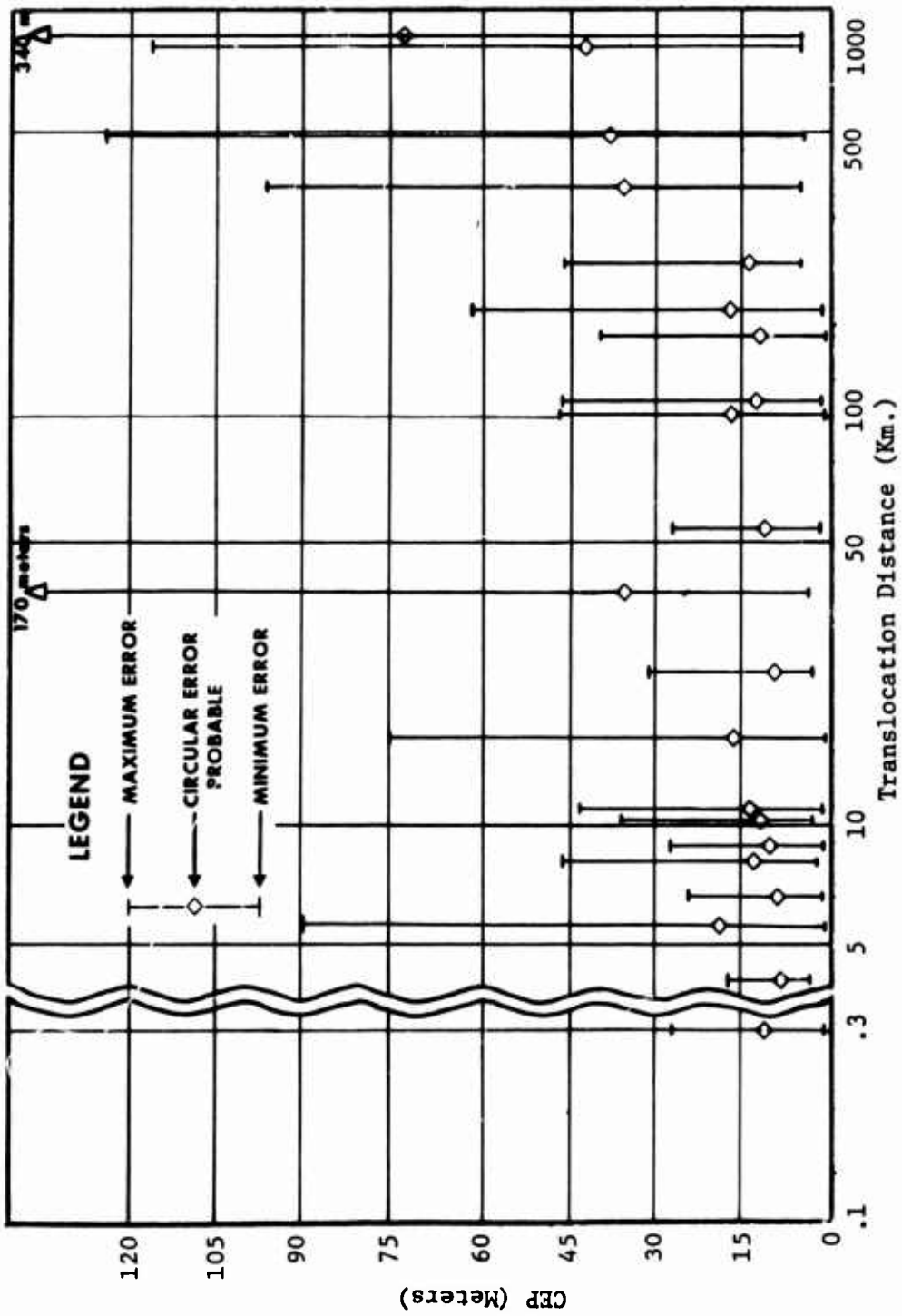
5. Any new translocation equipment development should include improved time recovery, shorter doppler counts, and three-dimension data reduction. The need for two-frequency operation must be evaluated on the basis of maximum system range and accuracy requirements.

Acknowledgments.

Many people and organizations made important contributions to the translocation test program. I would especially like to recognize Dr. C. Swartz and his staff of the DMA Topographic Center for reducing the Geociever data, C. Bunch of USAETL for modifying the Backpacks and keeping them running, G. Schmeidel, C. Oliver, and M. Hawker, all of USAETL for helping collect the data, and R. Muniz of USAETL for helping write the translocation program and evaluate the data. The Geocievers were furnished and operated by the DMA Topographic Center and the Naval Oceanographic Office. General support was provided by APL.

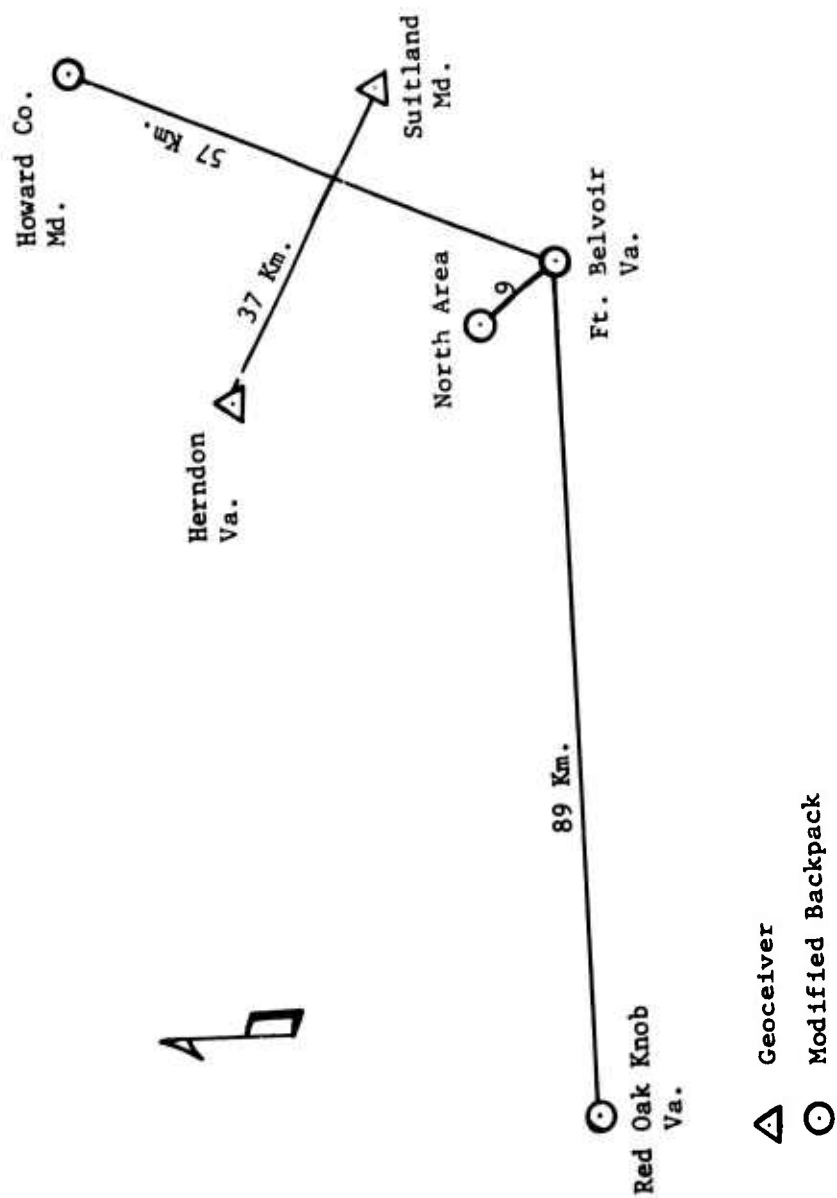
References.

1. Translocation Reference Manual, D. W. R. Denzler, APL Technical Memorandum TG 1174, January 1972.
2. "Military Potential Test of Manpack Navigation Satellite Locator System, AN/PRN-6 (XN-1) (Advanced Development Doppler Translocation Equipment)", Final Letter Report, Major Graham N. Babb, 5 June 1972.
3. "Present State of Navigation by Doppler Measurement from Near Earth Satellites", R. B. Kershner, APL Technical Digest, November-December 1965.
4. Program Requirements for Two-Minute Integrated Doppler Satellite Navigation Solution, J. B. Moffett, APL Technical Memorandum TG 819-1 (Rev), September 1971.



MILITARY POTENTIAL TEST OF TRANSLOCATION

Figure 1



STATION LOCATIONS AND TRANSLOCATION LINES

Figure 2

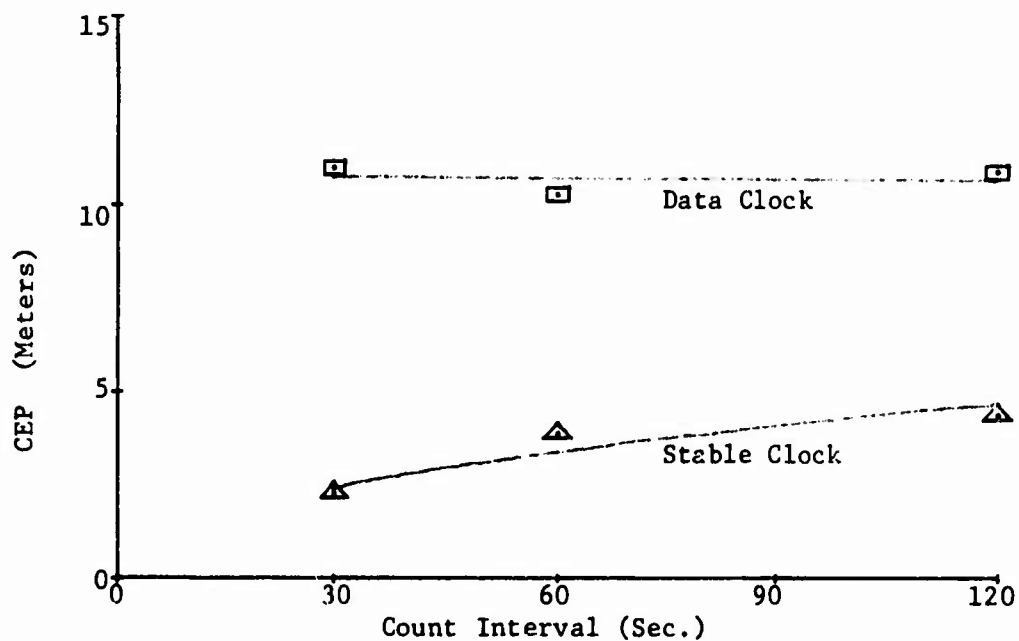


FIGURE 3 - 2-D GEOCEIVER ACCURACY

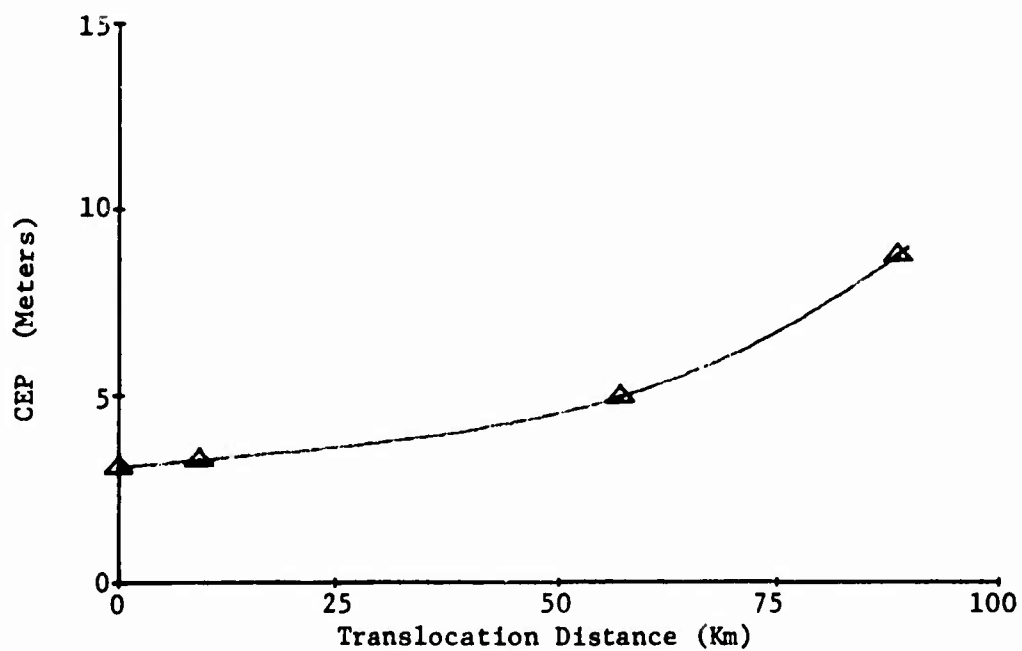
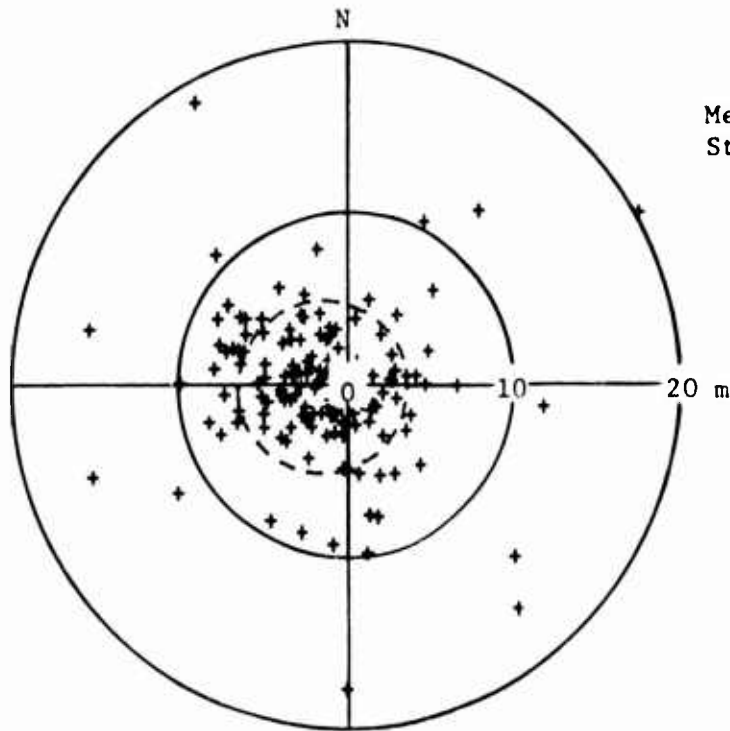


FIGURE 4 - 2-D MODIFIED BACKPACK ACCURACY

GLOECKLER



	N	E
Mean	-.3 m	-1.5 m
Std Dev	4.1	4.4

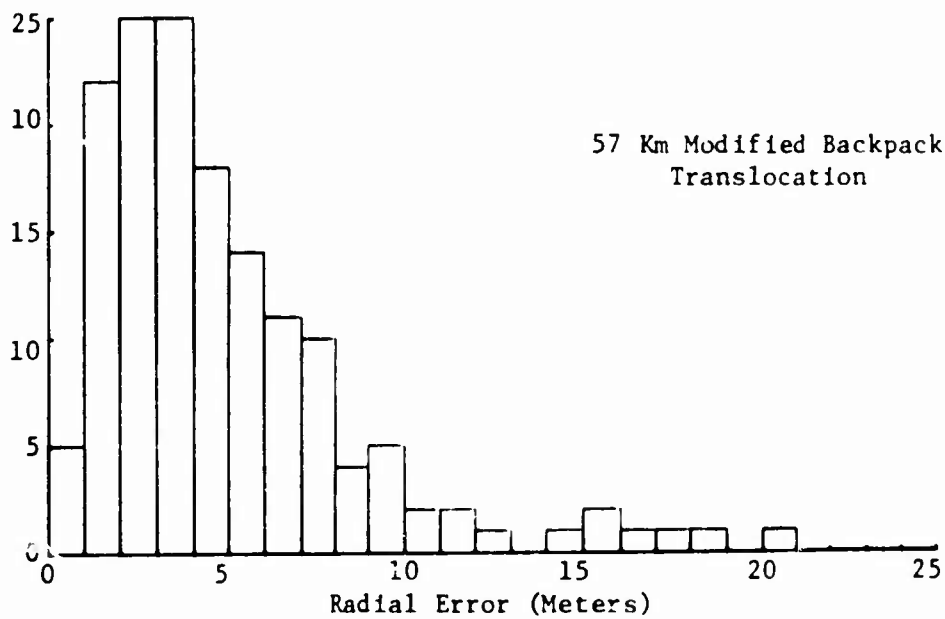
CEP 5.0 m

154 Passes

57 Km distance

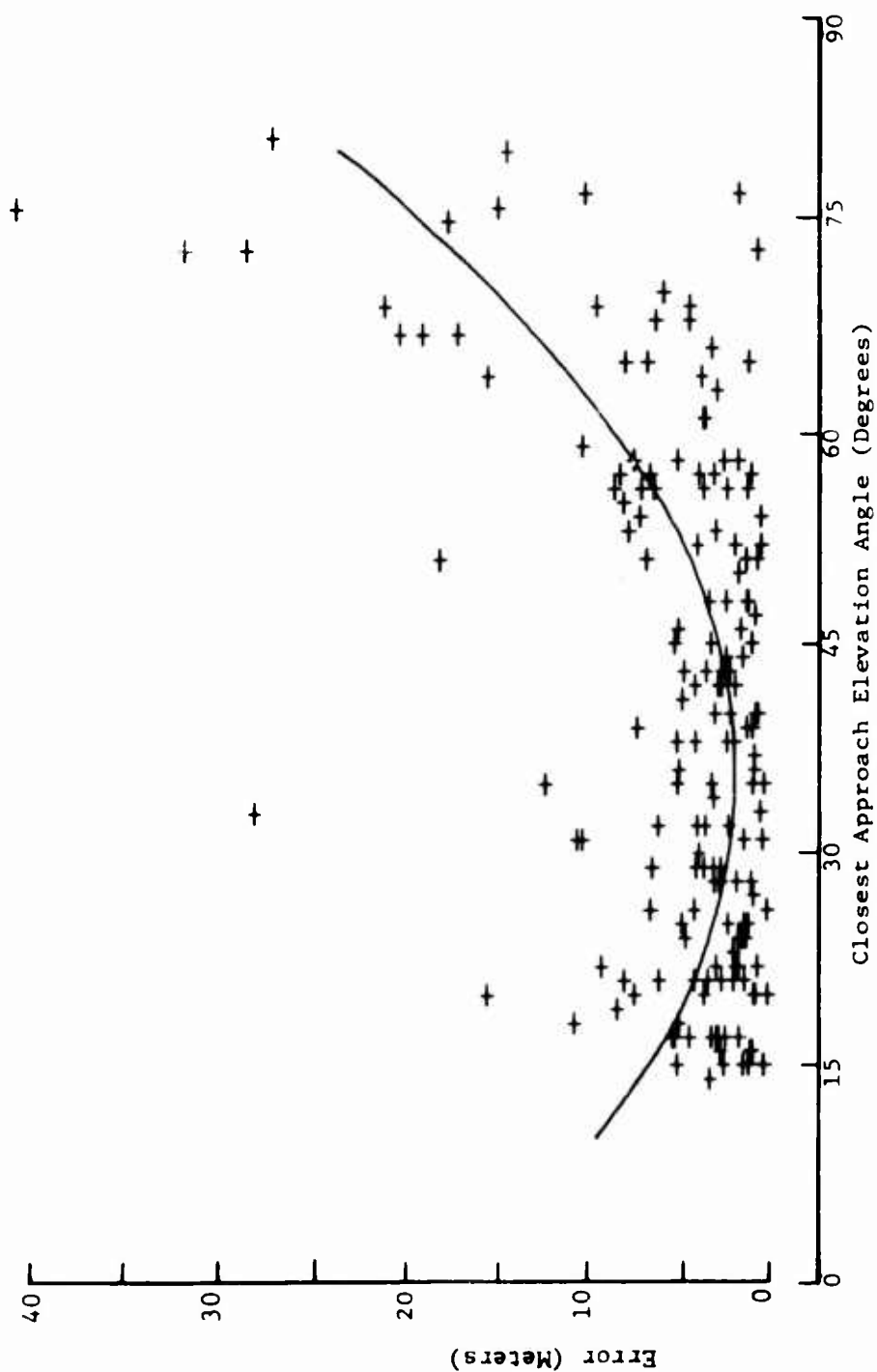
MODIFIED BACKPACK
TRANSLOCATION

Figure 5



57 Km Modified Backpack
Translocation

Figure 6 - RADIAL ERROR HISTOGRAM



LONGITUDE ERROR VERSUS ELEVATION ANGLE

Figure 7

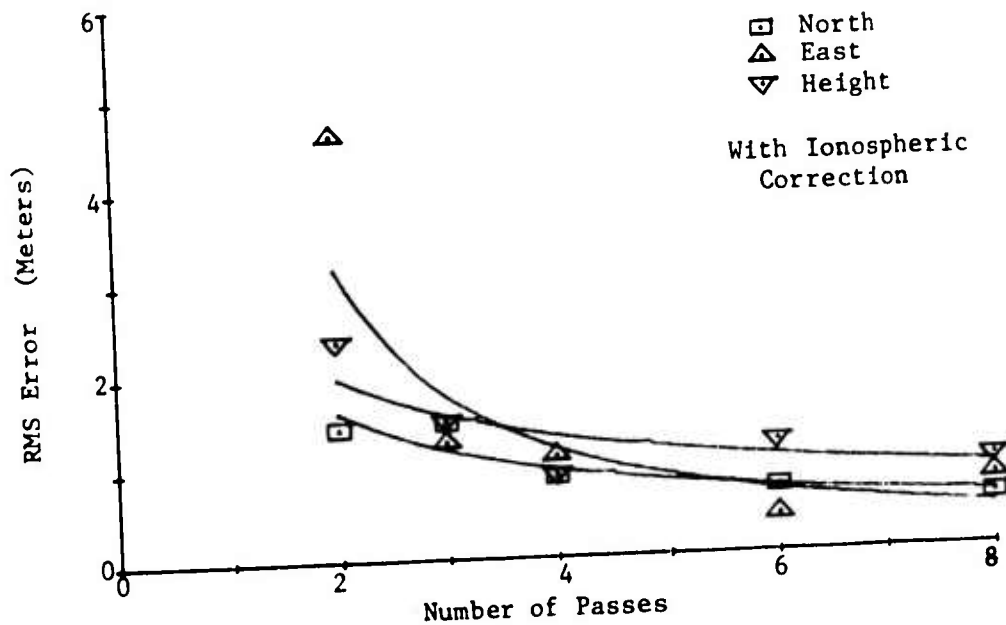


FIGURE 8 - 3-D TRANSLOCATION

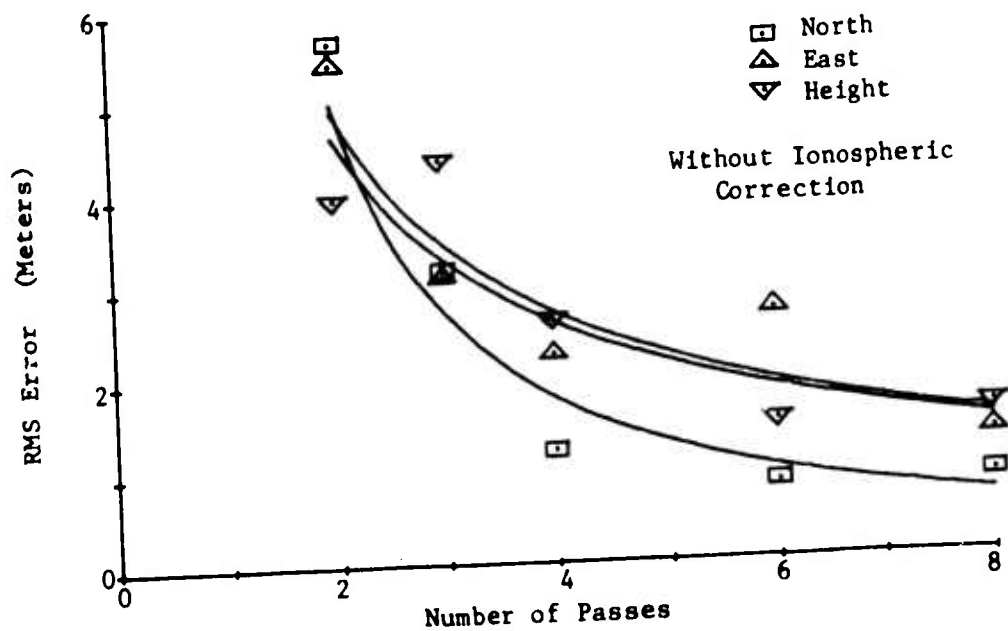


FIGURE 9 - 3-D TRANSLOCATION

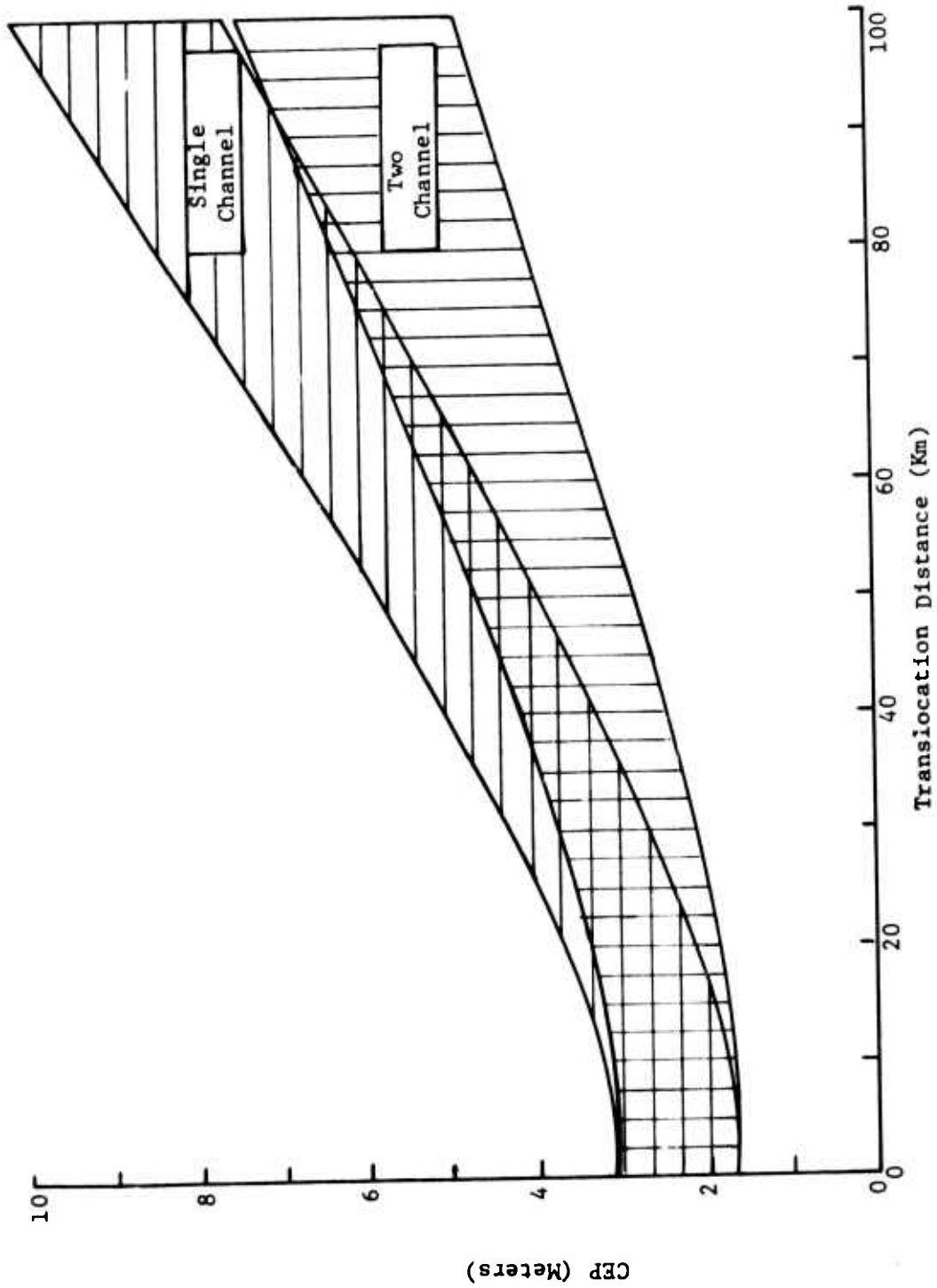


FIGURE 10 - PROJECTED TRANSLOCATION PERFORMANCE

ENVIRONMENT, CLOTHING AND PERSONAL
EQUIPMENT, AND MILITARY OPERATIONS

RALPH F. GOLDMAN Ph.D.
US ARMY RESEARCH INSTITUTE OF
ENVIRONMENTAL MEDICINE
NATICK, MA

INTRODUCTION:

Although the interaction between the psychological state of a soldier and his military performance is still not well defined, the interaction between his physiological status and his military performance is becoming much more predictable. In addition to working on such classic environmental injuries as heat stroke and frostbite, the U.S. Army Research Institute of Environmental Medicine also looks at: (a) such anthropometric features as stature, race and weight; (b) physiologic status with respect to physical work capacity, physical condition, age, state of acclimatization, hydration, sweat capacity and fatigue; (c) clothing and equipment with respect to their weight, thermal insulation (clo), the interference with sweat evaporation (i_m) and the design, fit and closures (pumping coefficients); (d) the physiological demands imposed by the assigned mission and its duration; and finally, (e) the terrain, ambient temperature, humidity, wind, precipitation and solar load of the environment. The interaction of these five types of factors suggests whether or not the soldiers in a unit will be physically able to complete a scheduled mission on time, with or without discomfort or difficulty, or will be so plagued by physical exhaustion, heat exhaustion or cold that the mission cannot be accomplished unless special precautions are taken or the scheduling of the mission is altered.

This paper will present a model for analysis of these interactions, will identify some specific environmental medical factors contributing to mission accomplishment problems, and will indicate the possibilities of a trade-off analysis of various solutions. Psychological research studies are being carried out on the interaction between environmental stress, physiological strain and mission perform-

ance, motivation and fatigue, but are not yet sufficiently advanced for incorporation.

The model presented does however provide a framework for identifying necessary additional human physiological and psychological studies, and for integrating new research findings to the solution of practical military problems.

Since, like most predictive models, the responses are predicted for a defined population, it is difficult to predict the response of a given individual with any great reliability, although the probable factors which must be considered in an individual case are identified. However, the most fruitful use of such a model is in defining:

(a) for the TACTICIAN - the physiological limits of his men, given their clothing, equipment, and the probable environment for a proposed mission.

(b) for the SURGEON - the probabilities of occurrence of environmental injuries (freezing and non-freezing cold injuries, heat exhaustion or stroke, physical exhaustion) and possible approaches to their prevention.

(c) for the CLOTHING AND EQUIPMENT DEVELOPER - the protection provided the soldier by his product, as well as the burden it imposes and the benefits of any proposed changes.

(d) for INDUSTRY, LABOR and HEW SPECIALISTS - the range of safe and/or comfortable environmental working conditions (this military relevant information is, in many areas, either all that is necessary, or all that may be available in any organized format).

METHODS:

Essentially the model (as shown in Figure 1) is a systems analysis of the stress imposed upon the soldier by his physical environment, clothing, equipment and mission. The model compares these stress demands with the physiological resources established by body size, age and sex, and modified by fitness, relative work load and acclimatization status, and also (although at present in very limited degree) the ability to mobilize these resources as modified by psychological factors. Finally, the model assesses the impact of any imbalance, in terms of such physiological strain responses as altered sweat production, hydration, heat production, body temperatures, heat content, and heart rate.

The model essentially looks at the classic physiologic heat balance equation of the human body and calculates the soldier's metabolic heat production (M) while performing a given mission (4,9), the solar heat load (R_s) if any (1), and the radiant and convective heat exchanges ($R+C$) allowed by the difference between his mean average skin temperature (T_s) and environmental temperature (T_A) (14). The net body heat balance may be negative as a result of these

$M+R \pm (R+C)$ exchanges (e.g., in a cool environment with inadequate clothing or low work levels), in which case falling deep body temperature (T_{re}) leads to shivering (20) and whole body hypothermia (2) or to cooling of extremity skin temperatures (T_{finger}), etc., to critical levels (23). Such heat losses can be predicted as a function of environmental temperature, clothing (including handwear, headwear, and footwear) and the mission activity level (18). Such problems occur because: (a) the capacity of the body to produce heat by shivering is limited (20); (b) truly adequate clothing insulation (clo) for extreme cold is unacceptably bulky and restrictive (16), and (c) the heat flow from the deep body heat stores to the extremities is drastically constricted as central temperature (T_{re}) is maintained to the detriment of extremity temperatures (T_{finger} , etc.) (12).

If, on the other hand, the balance of $M+R \pm (R+C)$ calculates out to a positive quantity, this represents the required evaporative cooling (E_{req}) (5) which the body must attempt to eliminate by evaporation of sweat (15). While man does tolerate heat stress better than cold, i.e., regulates body heat content much more efficiently in warm environments (11), the maximum evaporative cooling capacity (E_{max}) is limited (5) because: (a) the capacity of the body to produce sweat is limited (15); (b) any clothing, and especially such military items as body armor or NBCW protective clothing, has a limited permeability (i_m) which interferes with sweat evaporation (8,10) and (c) at high ambient humidities (ϕ_A) the difference between the vapor pressure of sweat at skin temperature (P_s) and ambient vapor pressure ($\phi_A P_A$) may not be enough to allow all the sweat produced, and allowed to transfer by the clothing, to evaporate (5). The unevaporated sweat will soak the skin and clothing, leading to the discomfort of a damp skin (13) or drip off the body without providing any cooling, uselessly depleting body water resources and thus contributing to the problem of dehydration (25) and heat exhaustion as cardiovascular stress become excessive (6) while body heat storage, and thus deep body temperature continues to rise toward heat stroke levels (15,19).

The heat imbalances predicted from the heat balance equation are expressed in kilocalories. An analysis of the approximate altered body heat storage (ΔS) of men during maneuvers, or carefully controlled chamber exposures, can be carried out by calculating a mean body temperature (T_b), assuming 2/3 of the total body weight (W) is represented by deep body temperature (T_{re}) and 1/3 by mean skin temperature (T_s) (14). Starting from a comfortable state of T_{re} at 98.6°F (37°C) and T_s at 92°F (33°C), the calculated mean body temperature (T_b) is 35°C. Taking the specific heat of the body as 0.83 kcal/kg°C, for a man weighing 70 kg each °C change in mean body temperature represents a change of 58 kcal (i.e., 0.83×70). Thus measured changes in body temperature can be directly related to the change in heat storage of the body, and the predicted imbalances in the heat

balance equation can be related to changes in body temperature (13).

RESULTS:

The model consists of a number of blocks, many already validated and published, some partially validated and others still being refined. A list of the input parameters that must be supplied (identified by ?) and calculated inputs (identified by !) is given in Table I; references for these input parameters are also included in Table I. Essentially, the inputs consist of factors (a) describing the soldier in characteristics from which his physiological capacities can be estimated, (b) factors describing his load, march rate, the terrain and grade (or other mission descriptors) from which the work demands and his heat production can be estimated, and finally (c) factors for his clothing insulation (clo), vapor permeability (i_m) and how these change with air or subject movement (pumping coefficient). The heat losses by radiation and convection, and the maximum evaporative heat loss, can be calculated from these three clothing parameters (clo , i_m and pumping coefficient), as a function of the ambient temperature, vapor pressure and wind. Subroutines are available for solar heat load (1), if any, and dehydration (25) (although not yet fully validated) and are being evaluated for extremity cooling (23) and developed for water immersion heat loss (3).

The model calculates a theoretical equilibrium state (5) at which, by adjusting deep body and skin temperatures, and shivering heat production or sweat production, the heat balance equation of the body can be satisfied; i.e., there would be no additional changes in heat storage or debt. Such equilibrium levels may be unattainable, or lethal, but the body will strain to achieve them. The outputs of the model describing these strains, are tabulated in Table II. A large number of validating and ad-hoc studies examining these input-output relationships have led to empiric derivation of the necessary time constants and delay times for the various physiologic responses to specific stress inputs; at present, the total response pattern for heart rate (16) and deep body temperature (5) as a function of time can be graphed with considerable confidence, both for acclimatized and unacclimatized men (7). Similar factors are being validated for extremity cooling (23) and dehydration (25).

Table III lists the critical levels for ambient environmental conditions, mean skin, finger and deep body temperatures, altered body heat content (storage or debt), dehydration, heat production and heart rate. When possible, a critical threshold is specified for levels of comfort, discomfort, performance decrement, tolerance and actual damage.

The combination of these input factors in Table I, output factors in Table II and critical thresholds in Table III allows the model to flag the times of occurrence of these critical states;

Table IV lists the various flags used, in four categories: comfort related; work related; heat related; cold related.

DISCUSSION:

Any prediction model, although perhaps theoretically useful and interesting, is only of limited use until validated across a range of military populations and until some estimate of the standard error of prediction is available. The prediction of heat production (4,9, 22) has been thoroughly validated across a range of military and civilian populations and appears to be good to $\pm 10\%$, which is about the usual precision of its measurement in the field. The prediction of troop mobility, i.e., distance covered per unit time or time to reach a given destination as a function of load (22) and terrain (24) has been validated on average soldiers (26) but requires more validation.

Prediction of deep body temperature in the heat as a function of ambient hot-wet or hot-dry environmental conditions (17), standard military uniforms (5), body armor (10,17) and NBCW protective clothing and mission activity (14,21) is also well validated and appears accurate to $\pm 0.2^\circ\text{C}$ (5) for a number of the combat arms, but may need some adjustment for elite military units. Prediction of heart rate for these situations (5) is also well developed, but for high heart rates the effects of physical condition and age need more work; nevertheless, the prediction for the standard volunteer military subjects available in the laboratory has a standard error of estimate of 6 to 10 beats per minute (15). More work is planned to characterize other military populations.

The effects of heat acclimatization are well defined and appear valid enough for practical use (7) although a wide range of hot-wet as well as hot-dry conditions has not been studied. The solar load prediction (1) also seems quite valid, but the effects on the soldier of combined metabolic and solar heat loads require more work. The effects of dehydration are just beginning to be incorporated into the model (25). Nevertheless, the overall heat tolerance prediction capability of the model has been tested in control field studies and a number of large scale maneuvers (the METOXE series and 69-10) and has proven to be quite practical for preventing heat casualties and establishing appropriate work-rest cycles. The trade-off analyses allowed by the model have also been extremely valuable during the development phases of new uniforms (21) and protective clothing and equipment (17).

On the cold side, the whole body cooling prediction model has been well validated in limited environmental chamber studies of sleeping bags and also in one field study. The extremity cooling model has proven most useful in design of handwear and footwear, and in prediction of the "worst case" cooling responses, but individual variability in the pattern of spontaneous rewarming (Cold Induced

GOLDMAN

Vaso Dilatation) severely complicates absolute prediction of extremity temperature (18). Since, however, such "worst case" prediction is generally applicable to a soldier who is thoroughly chilled or frightened it is still quite useful.

Current research emphasis is on improving the accuracy of the mobility and heart rate prediction to include various levels of physical fitness, age and motivation, extending the validation of the heat stress elements to include solar load and a wider range of dehydration conditions and, finally, developing more information to validly apply extremity cooling and whole body cooling prediction to field operations in the cold.

CONCLUSIONS:

The model presents the interaction of multidisciplinary factors drawn from (a) the theoretical physics of heat transfer, (b) the biophysics of clothing, (c) the physiology of metabolic heat production, distribution and elimination, and (d) related meteorological considerations. The model is extremely useful now for clothing and equipment developers since its rank ordering of the effects of their products has proven to be extremely accurate. In its present state, it is a useful guide for tactical planners, albeit a conservative one for two reasons; it has not been studied over a broad cross section of military units in the laboratory and, although its validation in field operations has included many of the combat arms of the Army and a few studies on Marines, the lack of motivation or realism common to such non-combat operations leads to a conservative estimate of performance capability on the one hand, while overlooking the degradation of performance that can be induced by the fear and pressures of actual combat. The model does provide the surgeon with a useful prediction of the relative risks of thermal environmental injury, the factors contributing to such risk and their relative proportions, but such prediction is for groups of subjects, not for individuals. It is also useful for engineers designing environmental conditioning equipment, for energy conservation planners and for developing industrial work regulations as proposed under the recent OSHA legislation. Finally, the model can perhaps be most useful to the scientist who developed it and to the Army Research Institute of Environmental Medicine in: (a) evaluating what additional research is needed; (b) deciding whether a proposed research study will provide militarily relevant information, and whether the environmental conditions selected are sufficiently strenuous to obtain measurable physiological differences between the parameters being studied, but not too strenuous to permit collection of meaningful data without undue risk to the subjects; and (c) in providing expert guidance and consultation to a wide range of military and civilian agencies.

References

1. Breckenridge, J.R. and R.F. Goldman. Solar Heat Load in Man. J. Appl. Physiol. 31: 659-663, 1971.
2. Gee, Gin K. and R.F. Goldman. Prediction of Whole Body Cooling. Biophysical J. 13: 213a, 1973.
3. Gee, Gin K. and R.F. Goldman. Heat Loss of Man in Total Water Immersion. The Physiologist 16: 318, 1973.
4. Givoni, B. and R.F. Goldman. Predicting Metabolic Energy Cost. J. Appl. Physiol. 30: 429-433, 1971.
5. Givoni, B. and R.F. Goldman. Predicting Rectal Temperature Response to Work, Environment and Clothing. J. Appl. Physiol. 32: 812-822, 1972
6. Givoni, B. and R.F. Goldman. Predicting Heart Rate Response to Work, Environment and Clothing. J. Appl. Physiol. 34: 201-204,
7. Givoni, B. and R.F. Goldman. Predicting Effects of Heat Acclimatization on Heart Rate and Rectal Temperature. J. Appl. Physiol. 35: 875-879, 1974.
8. Goldman, R.F. Tolerance Time for Work in the Heat When Wearing CBR Protective Clothing. Mil. Medicine. 128: 776-786, 1963.
9. Goldman, R.F. Energy Expenditure of Soldiers Performing Combat Type Activities. Ergonomics. 8: 321-327, 1965.
10. Goldman, R.F. Physiological Costs of Body Armor. Mil. Med. 134: 204-210, 1969.
11. Goldman, R.F. Environmental Limits, Their Prescription and Proscription. Intl. J. Environ. Stu. 5: 193-204, 1973.
12. Goldman, R.F., R.W. Newman and O. Wilson. Effects of Alcohol, Hot Drinks, or Smoking on Hand and Foot Heat Loss. Acta Physiol. Scand. 87: 498-506, 1973.
13. Goldman, R.F. Comfort, A Non-Subjective Approach. In: Proc. Kansas State University Visiting Lecturers. 1973.
14. Goldman, R.F. Quantitative Prediction of Physiological and Psychological Effects of Thermal Environment in Man. Archives des Sciences Physiologiques. 27: 73-78, 1973.
15. Goldman, R.F., E.B. Green and P.F. Iampietro. Tolerance of Hot, Wet Environments by Resting Men. J. Appl. Physiol. 20: 271-277, 1965.
16. Goldman, R.F. and A. Teitlebaum. Increased Energy Cost with Multiple Clothing Layers. J. Appl. Physiol. 32: 743-744, 1972.
17. Haisman, M.F. and R.F. Goldman. Physiological Evaluation of Armored Vests in Hot-Wet and Hot-Dry Climates. Ergonomics 16: in press, 1974.
18. Hanson, H.E. and R.F. Goldman. Cold Injury in Man: A Review of its Etiology and Discussion of its Prediction. Mil. Med. 134: 1307-1316, 1969.

GOLDMAN

19. Iampietro, P.F. and R.F. Goldman. Tolerance of Men Working in Hot Humid Environments. J. Appl. Physiol. 20: 73-76, 1965.
20. Iampietro, P.F., J.A. Vaughan, R.F. Goldman, M.B. Krieder, F. Masucci and D.E. Bass. Heat Production from Shivering. J. Appl. Physiol. 15: 632-634, 1960.
21. Joy, R.J.T. and R.F. Goldman. A Method of Relating Physiology and Military Performance: A Study of Some Effects of Vapor Barrier Clothing in Hot Climate. Mil. Med. 133: 458-470, 1968.
22. Kennedy, S.J., Goldman, R.F. and J. Slauta. The Carrying of Loads within an Infantry Company. US Army Natick Laboratories, Rpt C&PLSEL 108, May 1973.
23. Molnar, G., O. Wilson and R.F. Goldman. Analysis of Events Leading to Frostbite. Intern. J. Biometeorol. 16: 247-258, 1972.
24. Soule, R.G. and R.F. Goldman. Terrain Coefficients for Energy Cost Prediction. J. Appl. Physiol. 32: 706-708, 1972.
25. Soule, R.G. and R.F. Goldman. Dehydration and Exercise Performance in the Heat. J. Sports Med. in press, 1974.
26. Soule, R.G. and C.K. Levy (R.F. Goldman). Voluntary March Rate over Natural Terrains. Fed. Proc. 31: abs., 1972.

FIGURE 1 Diagram of a systems analysis of the stress inputs and physiological strain output responses of "the soldier"

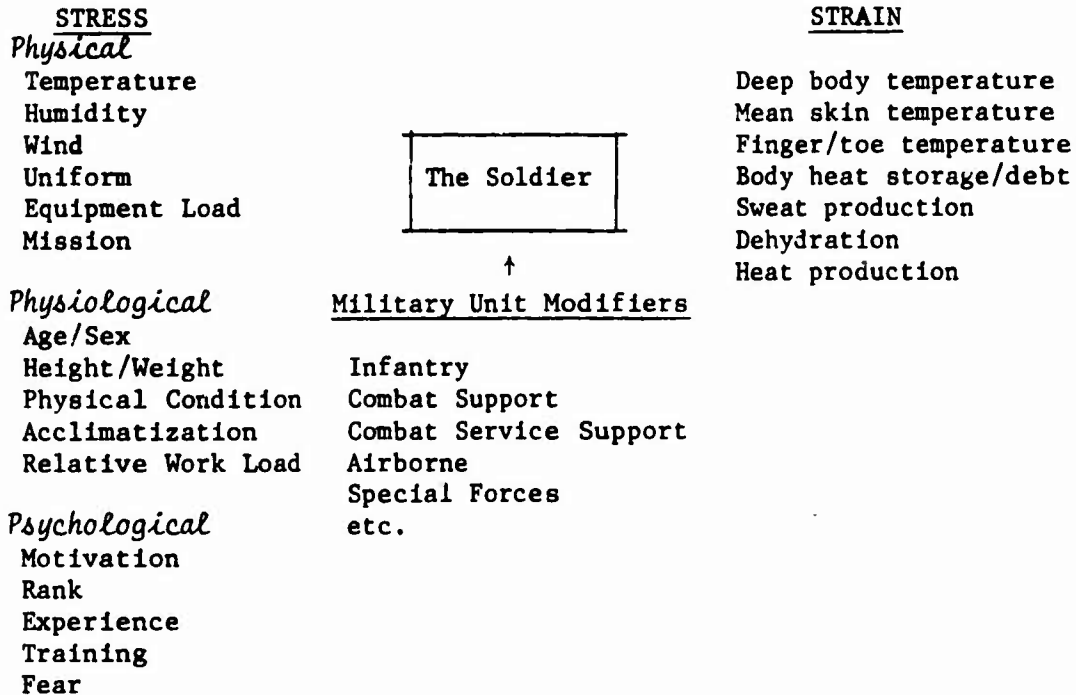


TABLE I Input Parameters

<u>Input #</u>	<u>Reference</u>
1 Height?	-
2 Weight?	-
3 + Sex?	-
4 Calculated surface area!	-
5 Effective area (heat rash, gas mask, etc)?	-
6 Load carried (clothing & equipment)?	12, 22
7 + Load carriage factor (any inefficiencies)?	15, 22
8 Terrain factor?	24
9 % Grade?	4
10 Specified march rate?	4
11 Predicted (march rate! <u>or</u> heat production)?	4, 22, 26
12 Uniform insulation (clo)?	1, 5, 10, 17
13 Uniform vapor permeability (i_m/clo)?	5, 10, 13, 17
14 Air or subject motion modifier?	5, 13, 14, 17
15 - Factor for sweat cooling efficiency (λ)?	5
16 + Handwear, footwear, headwear insulation?	13, 14, 18
17 Initial core temperature?	-
18 Estimated final skin temperature?	5
19 + Calculated core-skin conductance!	in preparation
20 Initial heart rate?	6, 7
21 + Age?	-
22 + Physical condition?	7, 14, 22
23 - Maximum work capacity!	6
24 - Motivation factor?	11, 22
25 Acclimatization status?	7
26 + Initial hydration status?	25
27 + Available drinking water?	22
28 + Extremity of concern?	18
29 Air temperature?	11
30 Relative humidity?	11
31 Wind speed?	5, 11
32 + Solar heat load subroutine?	1
33 - Water immersion subroutine?	3
34 + Mission activity?	4, 9
35 Duration of rest-work-recovery?	5, 6
36 Time interval of interest?	-

*Note: a + indicates information available for programming now but more validation may be needed; a - indicates more research needed.

TABLE II Outputs

- A) At theoretical equilibrium (N.B. may be lethal):
- 1 Core temperature rise from work
 - 2 Core temperature change by non-evaporative heat transfer
 - 3 Core temperature rise from limited evaporation
 - 4 Equilibrium (final) core temperature
 - 5 Decrease if fully acclimatized
 - 6 Equilibrium (final) heart rate
 - 7 Required evaporative cooling
 - 8 - Sweat production required
 - 9 - Rate of dehydration
 - 10 Maximum evaporative potential
 - 11 % Sweat wetted skin surface
 - 12 Cooling power
 - 13 Effective uniform insulation
 - 14 Effective uniform vapor permeability
 - 15 - Extremity circulatory heat input
 - 16 - Extremity time constant
 - 17 - Equilibrium (final) extremity temperature
- B) Tabulated values as a function of time for:
- 1 Core temperature
 - 2 - Mean skin temperature
 - 3 Heart rate
 - 4 - Extremity temperature
- C) Graphic display of time pattern response of:
- 1 Core temperature
 - 2 Heart rate
 - 3 - Extremity temperature

*Note: a - indicates more work is required before output is reliable.

TABLE IV Flags for Critical Levels of Comfort, Performance
or Environmental or Work Stresses

	<u>CRITERION</u>
<u>COMFORT RELATED:</u>	
Uncomfortable sweating	> 20% wet skin
Threshold of work discomfort	HR \uparrow 30b/min
Threshold of heat stress discomfort	$\Delta S \sim +25$ kcal
Threshold of cold stress discomfort	$\Delta S \sim -25$ kcal
<u>WORK RELATED:</u>	
Exceeds maximum work capacity	> 17 kcal/min
Exceeds voluntary hard work level	> 7 kcal/min
Work endurance difficult beyond:	
(a) 8 hours	5-6 kcal/min
(b) 2-3 hours	7-8 kcal/min
(c) 1 hour	10 kcal/min
(d) 15 min	12-15 kcal/min
<u>HEAT RELATED:</u>	
Exceeds voluntary heat endurance	$\Delta S \sim +80$ kcal
25% heat exhaustion collapse risk	$\Delta S \sim +120$ kcal
50% heat exhaustion collapse risk	$\Delta S \sim +160$ kcal
75% heat exhaustion collapse risk	$\Delta S \sim +200$ kcal
Increasing heat stroke risk	Tre > 106°F
<u>COLD RELATED:</u>	
Loss of manual dexterity	fingers < 60°F
Onset of shivering	$\Delta S \sim -80$ kcal
Violent shivering	$\Delta S \sim -150$ kcal
Non-freezing cold injury risk	any skin \leq 40°F
Freezing cold injury risk	any skin < 31°F
Increasing hypothermia risk	Tre < 95°F

GOLDMAN

The opinions or assertions contained herein are the private views of the author and are not to be construed as official or as reflecting the views of the Department of the Army or the Department of Defense.

A TECHNIQUE FOR THE VALIDATION
OF VEHICLE MODELS USING THE ROAD SIMULATOR

JAMES W. GRANT, MR.
U.S. ARMY TANK-AUTOMOTIVE COMMAND
WARREN, MICHIGAN 48090

INTRODUCTION

One of the more important aspects of vehicle design is mechanical mobility. This term is a measure of how fast a vehicle can get from Point A to Point B without breaking and yet preserving the cargo and maintaining driver comfort. Mechanical mobility may be divided into five categories: Rough terrain mobility, soft terrain mobility, water mobility, vegetation mobility, and high speed mobility. Each of these five mobility divisions may be subdivided into ride comfort (including cargo), durability and handling. The TACOM road simulator has been used extensively to investigate the durability and ride characteristics of existing and proposed military vehicles relative to rough terrain and high speed mobility. The real payoff for simulation testing, however, lies not in the durability testing of production vehicles and subsystems but in using it as a tool to interface computer models with prototype hardware to evaluate subsystems prior to integration to form a prototype vehicle system; and finally the road simulator would be used to perform the initial performance and durability evaluation of the new vehicle prior to field tests.

The process from the vehicle concept to prototype field testing is an iterative process using the methodology proposed in this paper. The first step in the iteration is to formulate a mathematical model using the dynamic equations of motion for the suspension system and relevant components of a vehicle concept. This model is then subjected to the mission profile for which the vehicle is intended. The suspension parameters are then tuned to minimize a

GRANT

cost factor, which is a function of ride, handling, cargo comfort, speed, etcetera. The second step is to manufacture prototype suspension components having the parametric characteristics derived from the mathematical model. These components are then evaluated to determine their real parametric characteristics. The third step then is to plug the measured parameters into the mathematical model and reevaluate to determine the acceptability of any parametric changes. The fourth step is to build a prototype vehicle, install it on the road simulator and subject the vehicle system to the mission profile. The fifth step is the final adjustment of the mathematical model to match its response to the response of the vehicle on the road simulator. This model may now be used for any dynamic studies on the vehicle which precipitate from field tests.

This paper will be limited to a description of some of the techniques which have been used at TACOM within the Surface Mobility Division aimed at validating mathematical models.

CONCLUSIONS AND RECOMMENDATIONS

It is apparent from this study that although the simple model tracks the hardware fairly well, a more sophisticated model is needed. This study also shows that the actuator to vehicle interfaces and centering radius rods have a significant effect on the response of the vehicle. These effects should either be minimized during initial simulator design or should be included as inputs to the computer model. The techniques used for comparing responses are adequate except for the possible addition of computing a correlation coefficient. Some recommendations related to the mathematical model are:

1. Treat the frame as an elastic member.
2. Separate the following into separate mass-spring-damper systems.
 - a. Cab.
 - b. Engine.
 - c. Cargo Box.
3. Include roll in the simulation.
4. Increase the number of transducer locations and hence the number of correlation points.

GRANT

No ride optimization or parameter adjustment was attempted due to the oversimplification of the model which was a result of inadequate computer capacity. A computer with sufficient capacity to continue this study has been acquired by the Surface Mobility Division at TACOM. The XM705 program is complete so continued model validation should be performed using some other vehicle scheduled for evaluation on the road simulator.

BACKGROUND

The introduction to this paper has outlined a proposed methodology for the computer aided design and development of a vehicle from initial concept through the prototype phase and into field testing. This paper will focus specifically on the fifth step which is the comparison of the model response with the response of the vehicle system on the road simulator. The objective of this study is to initiate steps to fill that particular gap in the proposed methodology. Previous work within the Surface Mobility Division at TACOM has resulted in a continuous parameter tracking technique for determining spring rates and damping coefficients for non-linear suspension components which complements the classical direct reading techniques for measuring vehicle parameters. The work contained in this paper was initiated to supplement a durability test on the XM705 1½-Ton, 4X4, Cargo Truck to be performed using the road simulator. The XM705 is shown mounted to the road simulator in Figure 1.

The instrumentation required for the durability test was sufficient to allow this "first cut" mathematical model validation to be made. The mathematical model was, therefore, designed to utilize the existing test setup.

MODEL FORMULATION

The vehicle model was limited to four degrees of freedom in the pitch and bounce modes. At the time this study was initiated, sufficient analog computer capacity to include roll freedom was unavailable. Since this was to be the initial attempt at tracking vehicle responses to model responses in real time and since roll could be eliminated on the road simulator, the elimination of roll freedom in the model would not invalidate the technique. Further simplifying assumptions were as follows:

1. The sprung mass is modeled as one rigid body.
2. The pitch angle, θ , is assumed small such that $\sin\theta = \theta$ and $\cos\theta = 1$.

GRANT

The vehicle is coupled through the wheel spindles to the road simulator as shown by the arrows in Figure 1 thus eliminating the tires and unsprung masses from the system model. The model then may be represented as shown in Figure 2.

Equations of Motion.

The free body diagram of the sprung mass is shown in Figure 3.

Where:

\ddot{y}_{CG} = Vertical acceleration of the center of gravity.

$\ddot{\theta}$ = Pitch acceleration.

a = Distance from center of gravity to the center of the front suspension = 6.15 ft.

b = Distance from center of gravity to the center of the rear suspension = 5.10 ft.

F_{FT} = Total front suspension force.

F_{RT} = Total rear suspension force.

m = Sprung mass = 139.75 slugs.

I_{θ} = Sprung mass pitch inertia = 750 slug-ft²

The equations of motion for this system then are:

$$m\ddot{y}_{CG} = F_{FT} + F_{RT} - mg \quad (1)$$

$$I_{\theta}\ddot{\theta} = aF_{FT} - bF_{RT} \quad (2)$$

The total suspension force is a summation of the spring force, shock absorber force and frictional forces. The spring forces and shock absorber forces are obtained from curves shown in Figures 4, 5 and 6. The interleaf friction for the leaf springs was measured in the laboratory. This friction is proportional to the spring force. The constant of proportionality fell in the range 10 to 15 percent of the spring load, 10 percent for the rear springs with three leaves and 15 percent for the front springs with five leaves.

GRANT

Equations 1 and 2 may now be rewritten

$$m\ddot{y}_{CG} = F_{FS} + F_{FD} + F_{FF} + F_{RS} + F_{RD} + F_{RF} - mg \quad (3)$$

$$I_G \ddot{\theta} = a F_{FS} + a F_{FD} + a F_{FF} - b F_{RS} - b F_{RD} - b F_{RF} \quad (4)$$

Where F_{FS} and F_{RS} = front and rear spring forces

F_{FD} and F_{RD} = front and rear shock absorber forces

F_{FF} and F_{RF} = front and rear friction forces.

Also from the above discussion:

$$F_{FF} = .15 F_{FS}$$

and $F_{RF} = .10 F_{RS}$

Samples of the analog computer circuits required for the model simulation are shown in Figures 7 through 12.

ROAD SIMULATOR

The road simulator used in this study and shown in Figure 1, imparts vertical excitations to the wheel spindles of the vehicle using four linear electro-hydraulic servo controlled actuators. These actuators are capable of 200 inches per second linear velocity, have twelve inches of usable stroke and respond to 100 hertz with a double amplitude of .0005 inches.

As mentioned previously, the vehicle is attached through its wheel spindles to the actuators. The left front attachment is indicated by the arrow in Figure 1. The interface hardware allows six degrees of freedom at each wheel in order not to unrealistically load the axles and suspension. The vehicle is attached through the wheel spindles (tires and wheels removed) to provide a more positive attachment to the actuators which in turn prevents vehicle ejection from the road simulator due to high amplitude inputs. One major benefit, related to this paper, of removing the tire is that a reliable tire model is no longer required to validate the suspension model.

GRANT

SINE WAVE RESPONSE

The computer model and the road simulator were excited simultaneously to a sine wave of five inches or maximum attainable double amplitude. The frequency was increased from .5 hertz to 8 hertz. Figure 13 shows the responses of the model and the vehicle at their centers of gravity to these excitation signals. Several conclusions may be drawn from comparing the recordings and also by considering observations made during each of the test runs:

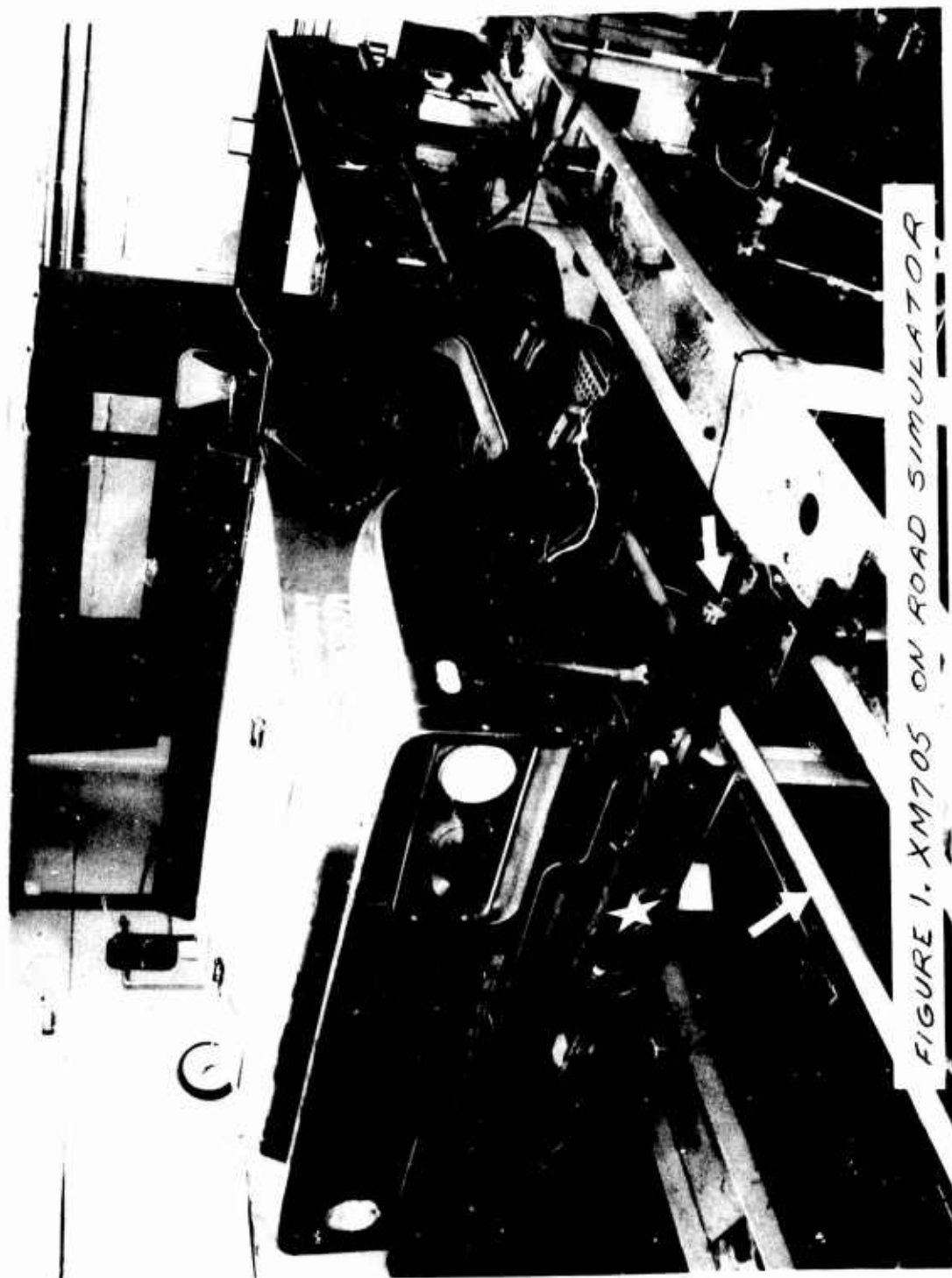
1. The noise superimposed on the acceleration signal from the road simulator is a result of mechanical noise (system clearance) in the interfaces between the spindles and the actuators.
2. The attenuation of the acceleration signal from the road simulator at three hertz was a result of the vehicle's center of gravity scribing an ellipse instead of a vertical straight line. This was caused by the inputs of the radius rods which were used to center the vehicle on the road simulator.
3. Overall, the model response compared well with the response of the vehicle on the road simulator.

The next step is to compare responses to terrain related random excitation signals.

RANDOM RESPONSE

The model and the vehicle on the road simulator were then subjected to a field vibration related random input. Maximum actuator excursion was twelve inches. The results of this random response are shown in Figures 14 and 15. Figure 14 shows the time histories of the vertical frame accelerations at a point centered above the front suspension. The vehicle response lags the model response due to hydraulic lag and filter lag. A low pass filter (40Hz) was inserted at the accelerometer output to clean up the signal for ease of presentation and comparison. Figure 15 shows the power spectral density curves computed from the time signals in Figure 14. The resonant points at two and five hertz agree well. The reduced amplitude of the vehicle response at the five hertz resonant point was caused by the effects of the radius rods as explained in the sine wave response discussion.

GRANT



GRANT

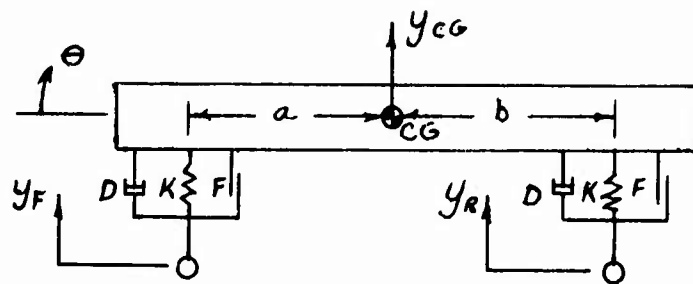


FIGURE 2. MODEL DIAGRAM

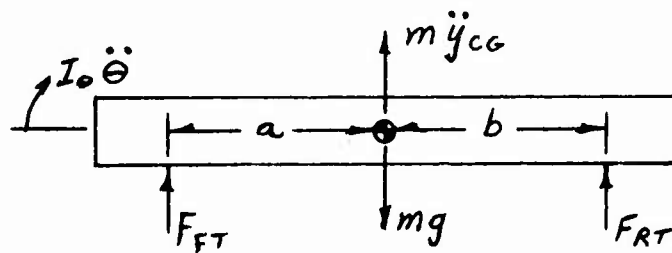


FIGURE 3. FREE BODY DIAGRAM

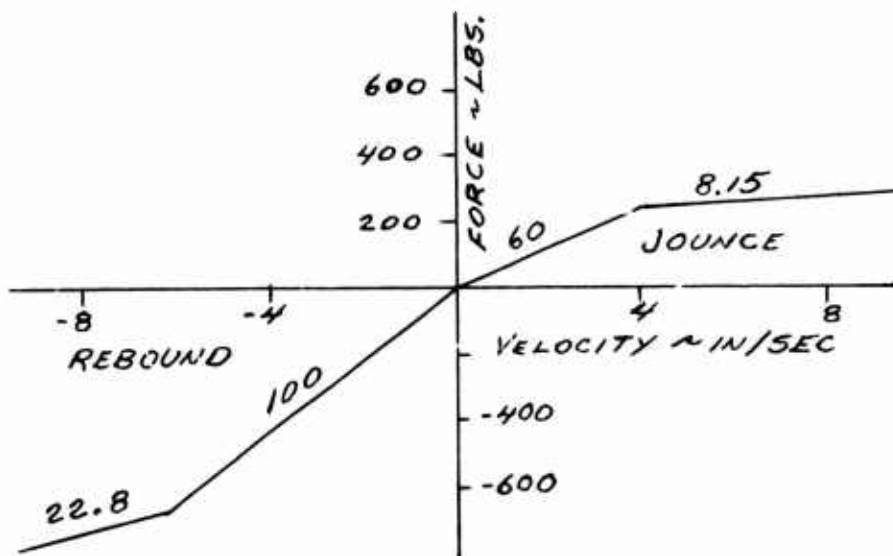
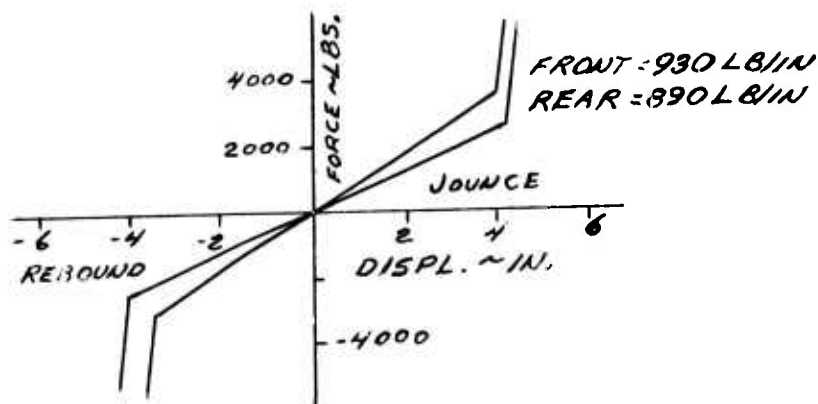


FIGURE 4. SHOCK ABSORBER CURVE



FIGURES 5&6. FRONT & REAR SPRING CURVES

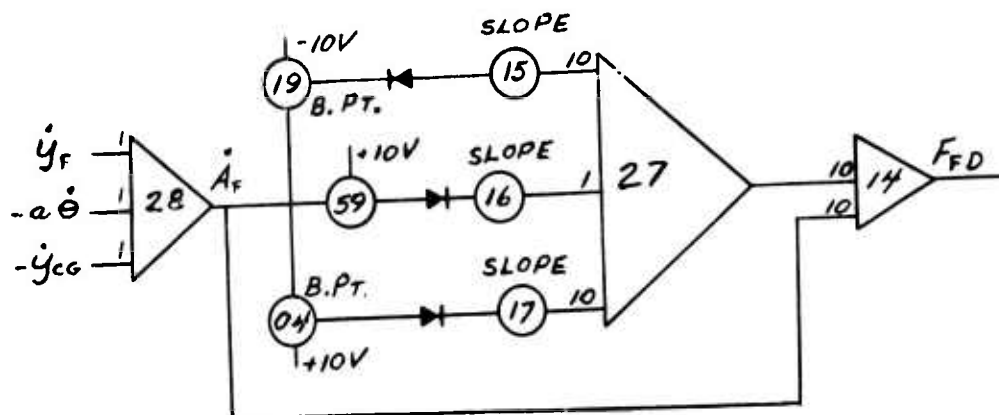


FIGURE 7. SHOCK ABSORBER ANALOG DIAGRAM

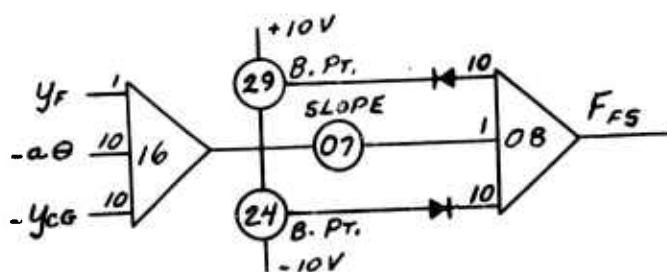
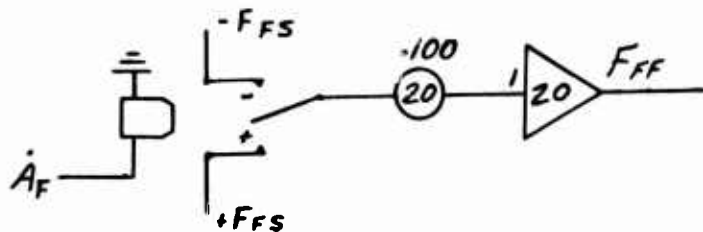
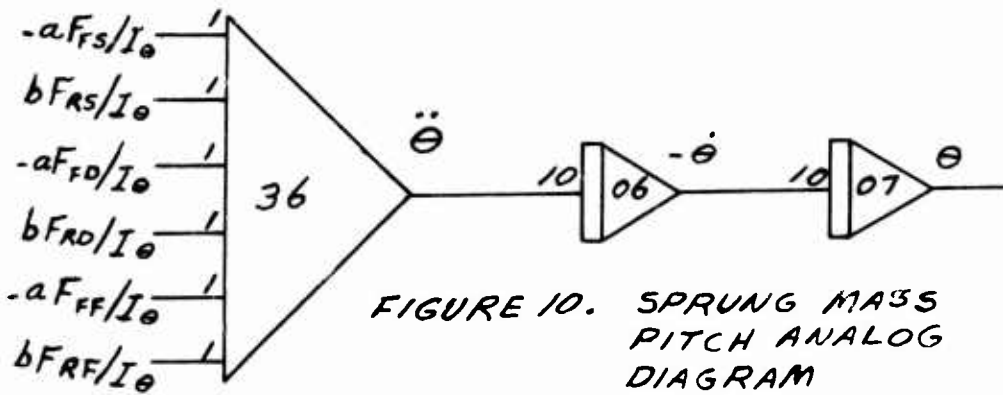
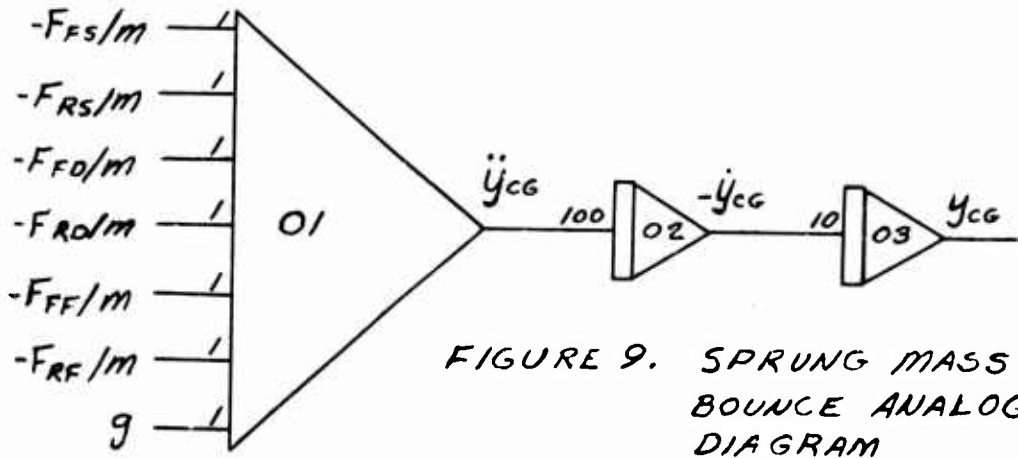


FIGURE 8. SPRING FORCE ANALOG DIAGRAM



GRANT

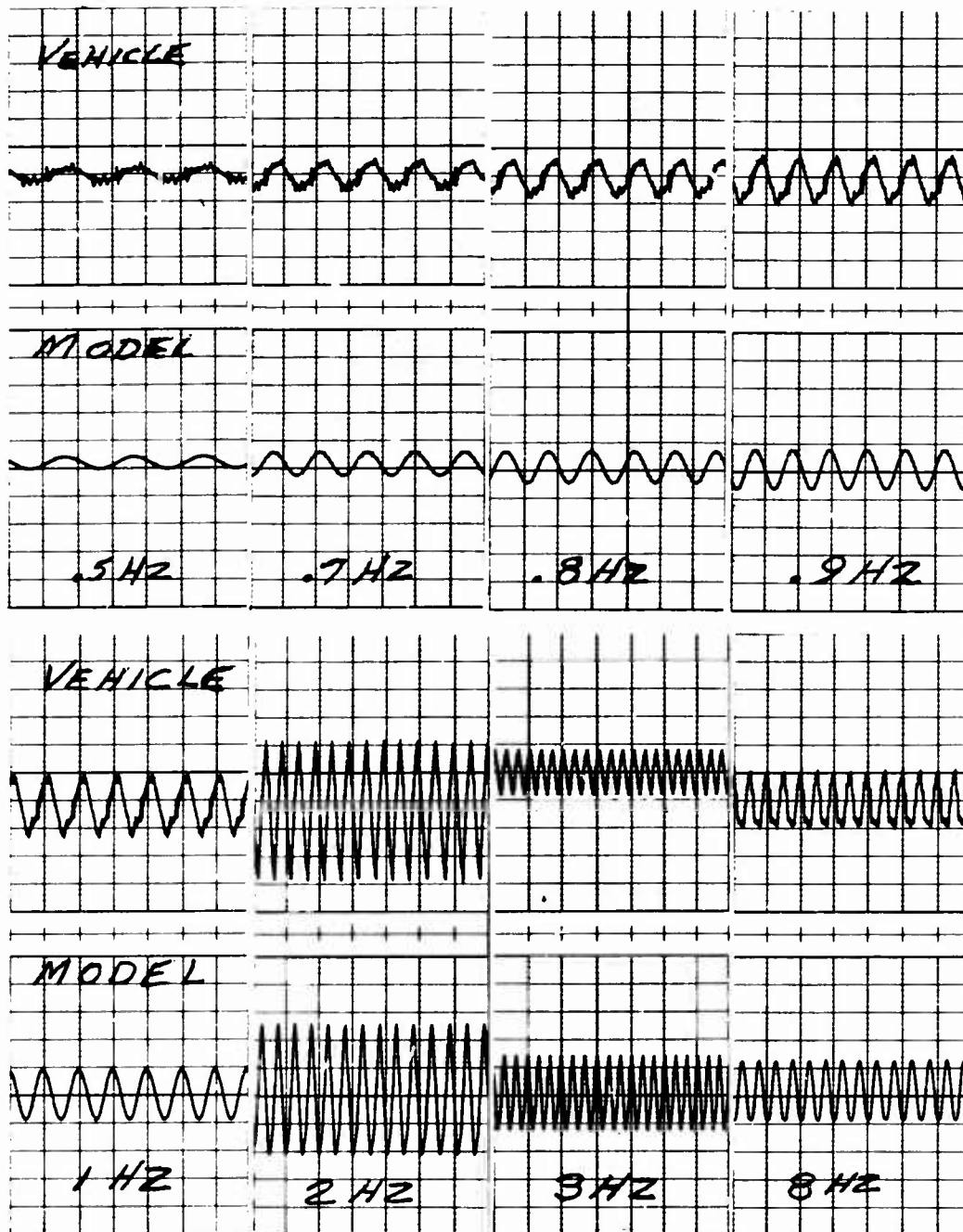


FIGURE 13. CENTER OF GRAVITY
ACCELERATION - 0.19/LINE

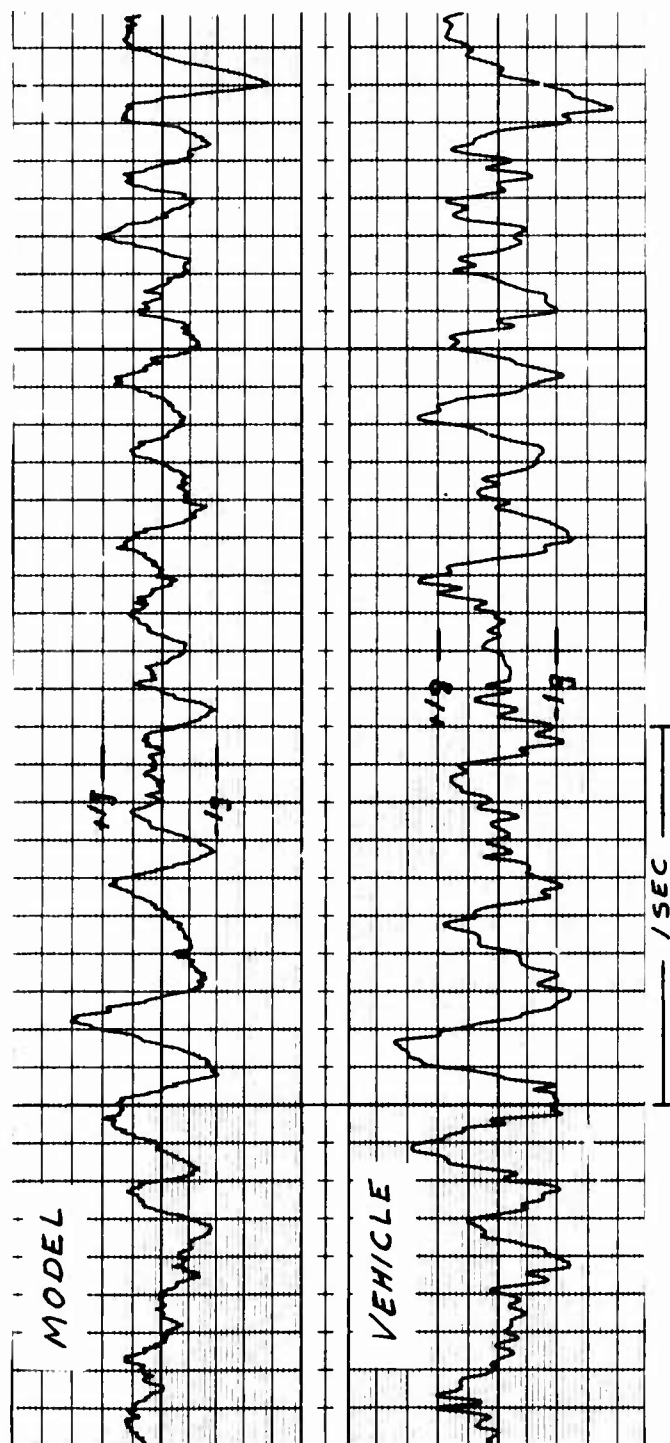


FIGURE 14. FRAME ACCELERATION RESPONSE

GRANT

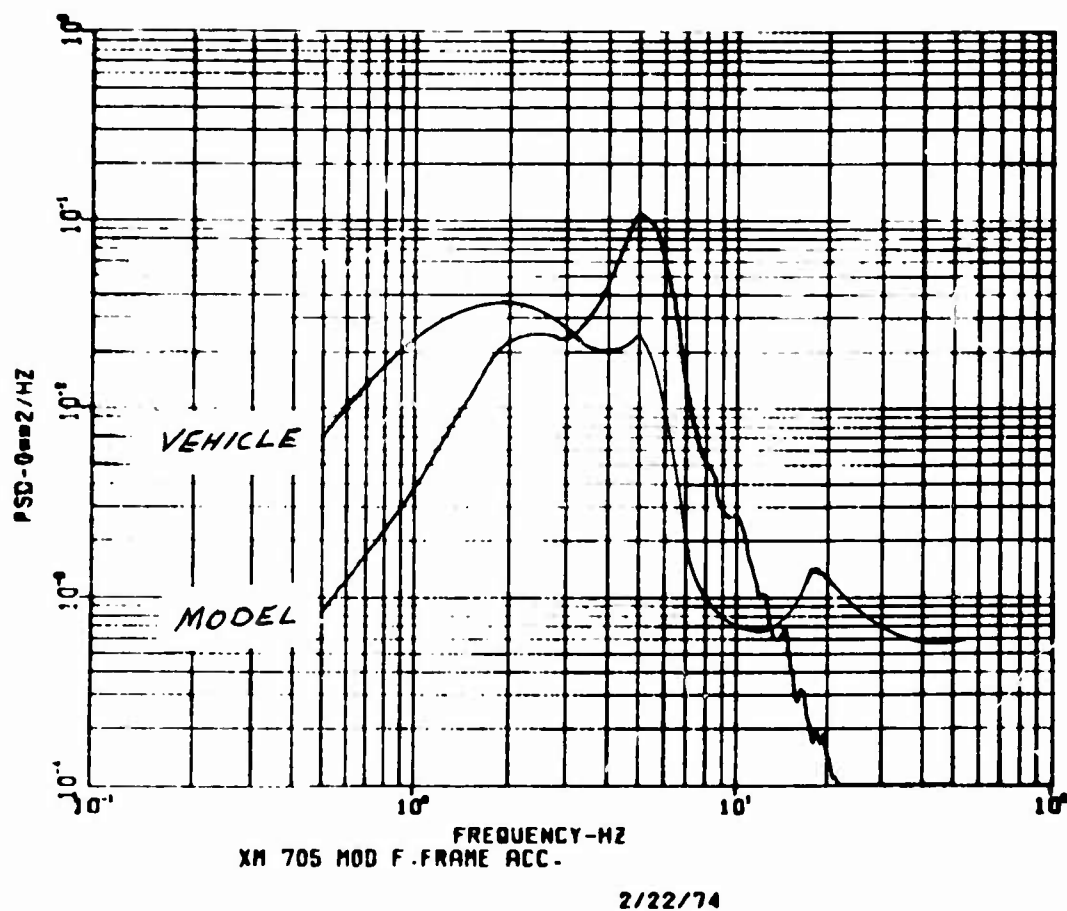


FIGURE 15. FRAME ACCELERATION PSD

HELICOPTER GROUND RESONANCE ANALYSIS
IN LIGHT OF ARMY REQUIREMENTS

CHARLES E. HAMMOND, Ph. D.
LANGLEY DIRECTORATE
UNITED STATES ARMY AIR MOBILITY R & D LABORATORY
HAMPTON, VIRGINIA

The Army has in recent years refined its methods of procuring aircraft. Until recently the Army had only procured aircraft which were more or less off-the-shelf items. That is, the aircraft were either developed for other military services or for civilian use and were adapted to meet Army needs. As a result, the Army was not significantly involved in the writing of specifications which affected the overall design of the vehicle. Operational experience has indicated, however, the need for specifications which reflect the particular requirements of Army aviation. This paper is concerned with requirements which have been imposed in the area of helicopter mechanical instability, or ground resonance as this phenomenon is commonly known, and the impact which these requirements have on the analyst.

The ground resonance problem has been with rotary wing aircraft since their inception, but the phenomenon was not well understood until many years after its first occurrence on the early autogyros. The instability was originally thought to be aeromechanical in nature, however, later analysis confirmed that the instability could be predicted based solely on mechanical considerations. The report by Coleman and Feingold (1) has become the classical reference on this problem. This classical treatment of the problem was expanded by Brooks (2) and Bielawa (3) to include additional degrees of freedom and to experimentally verify the analysis.

All these analyses showed that the ground resonance instability involved a mechanical coupling of the inplane degrees of freedom of the rotor blades with the rigid body degrees of freedom of the helicopter on its undercarriage. The analytical results also indicated that the instability could be eliminated within the operating

Preceding page blank

HAMMOND

rotor speed range of any particular helicopter by properly selecting (i) the damping and stiffness characteristics of the undercarriage and (ii) the inplane damping of the blades.

Because of operational considerations, the Army has imposed requirements that the helicopter not only be shown free from mechanical instability under normal operating conditions but freedom from instability must also be shown for a variety of abnormal conditions. Some of these abnormal conditions are: operation on ice, flat tire and flat strut on one side, operation from a 12° slope on any heading, and operation with one blade damper inoperative. All but the last of these requirements can be handled by appropriate modifications of the basic parameters in the classical analysis of Reference 1. As will be shown later, the requirement for stability with one blade damper inoperative causes one of the basic assumptions in the classical analysis to be violated and thus a new analysis method must be developed.

The investigation of the ground resonance problem when one blade damper is inoperative is the subject of this paper. A method will be presented for handling the problem and comparison will be made with two methods which have previously been used in attempting to satisfy the Army requirement. Instabilities which arise when one blade damper is inoperative will be examined and, finally, means for eliminating these instabilities will be discussed.

SYMBOLS

c_i	Lag damping rate
c_x	Effective hub damping in x-direction (longitudinal)
c_y	Effective hub damping in y-direction (lateral)
e	Lag hinge offset
I_b	Second mass moment of blade about lag hinge
k_i	Lag spring rate
k_x	Effective hub stiffness in x-direction (longitudinal)
k_y	Effective hub stiffness in y-direction (lateral)
m_b	Blade mass

HAMMOND

m_x	Effective hub mass in x-direction
m_y	Effective hub mass in y-direction
N	Number of blades in rotor
S_b	First mass moment of blade about lag hinge
t	Time
x_h, y_h	Coordinates of hub in fixed reference frame
ζ_i	Lag deflection of i^{th} blade
η_i	Defined by Equations (2)
v_o	Defined by Equations (2)
ψ_i	Azimuthal location of i^{th} blade
Ω	Rotor speed
ω_{o_i}	Defined by Equations (2)

MATHEMATICAL FORMULATION

A complete derivation of the equations of motion for the ground resonance problem is presented in Reference 4. This development will not be repeated here, but the equations together with the underlying assumptions will be given in order to discuss the implications of the one-blade-damper-inoperative requirement.

It is assumed, as is done in Reference 1, that the helicopter on its landing gear can be represented by effective parameters applied at the rotor hub. It is further assumed that only inplane motions of the hub and blades are important in determining the ground resonance characteristics of the helicopter. Thus the degrees of freedom to be considered consist of two inplane hub degrees of freedom and one lead-lag degree of freedom for each blade in the rotor. The mathematical model to be used in the analysis is shown in Figure 1. Note that in the figure only a typical blade is shown. The analysis is formulated for a rotor having N blades, and each blade is assumed to have a rotational spring and damper which act about the lag hinge. Further, it is assumed that each of the blades may have different lag spring and lag damper characteristics. This last assumption is

HAMMOND

necessary in order to be able to treat the one-blade-damper-inoperative situation and represents a major departure from the classical analysis. In References 1, 2, and 3 each blade was assumed to have identical properties and, as will be indicated later, this assumption leads to considerable simplification of the equations of motion. With respect to the hub degrees of freedom, it is assumed that, in the absence of the rotor, the longitudinal and lateral motions of the hub are uncoupled. This is an approximation, but it is an assumption made in Reference 1 and one generally used in mechanical stability analyses.

Based on the above assumptions, the equations of motion for the rotor-hub system may be written as

$$\left. \begin{aligned} \ddot{\zeta}_i + \eta_i \dot{\zeta}_i + \left(\omega_{o_i}^2 + \Omega^2 v_o^2 \right) \zeta_i &= \left(v_o^2 / e \right) \left[\ddot{x}_h \sin \psi_i - \ddot{y}_h \cos \psi_i \right] \\ i &= 1, 2, \dots, N \\ \\ (m_x + Nm_b) \ddot{x}_h + c_x \dot{x}_h + k_x x_h &= S_b \sum_{i=1}^N \left[\left(\ddot{\zeta}_i - \Omega^2 \zeta_i \right) \sin \psi_i \right. \\ &\quad \left. + 2\Omega \dot{\zeta}_i \cos \psi_i \right] \\ \\ (m_y + Nm_b) \ddot{y}_h + c_y \dot{y}_h + k_y y_h &= -S_b \sum_{i=1}^N \left[\left(\ddot{\zeta}_i - \Omega^2 \zeta_i \right) \cos \psi_i \right. \\ &\quad \left. - 2\Omega \dot{\zeta}_i \sin \psi_i \right] \end{aligned} \right\} (1)$$

where

$$\left. \begin{aligned} v_o^2 &= e S_b / I_b \\ \omega_{o_i}^2 &= k_i / I_b \\ \eta_i &= c_i / I_b \end{aligned} \right\} (2)$$

and small displacement assumptions have been made on ζ_i , x_h , and y_h in order to linearize the equations. The parameter ψ_i describes the azimuthal position of the i^{th} blade at time t and is given by

$$\psi_i = \Omega t + 2\pi(i - 1)/N \quad (3)$$

The equations of motion for the system thus consist of $(N + 2)$ coupled second-order differential equations in which the coupling terms have periodic coefficients. The periodic coefficients arise because the blade equations, the first N of Equations (1), are written in a rotating reference system whereas the hub equations, the last two of Equations (1), are written in a fixed system.

In order to eliminate the periodic coefficients in Equations (1) Coleman and Feingold (1) transformed the blade equations into the fixed reference system. This transformation, which greatly simplifies the equations of motion, is only possible if the rotor is isotropic with three or more blades. An alternate procedure is presented by Hammond (4) for eliminating the periodic coefficients if the rotor is nonisotropic, as is the case for one blade damper inoperative, but the hub is isotropic. This procedure involves transforming the hub equations of motion into the rotating frame of reference and requires that the rotor have two or more blades. Since the hubs of most, if not all, currently operational helicopters are nonisotropic this last procedure is only useful for determining the general nature of instabilities which occur when one blade damper is inoperative.

Thus for the general case of a nonisotropic rotor coupled with a nonisotropic hub one is faced with the problem of determining the stability of a system which is described by a set of second-order differential equations having periodic coefficients. It has been shown by Hammond (4) that the Floquet Transition Matrix method described by Peters and Hohenemser (5) and Hohenemser and Yin (6) provides an effective means for determining the stability characteristics of the system described by Equations (1). This method is essentially an eigenvalue method which is based on the Floquet-Liapunov theorem (7) for systems having periodic coefficients. Thus the stability characteristics of the system are direct outputs of the method.

Two methods used in the past for treating the one-blade-damper-inoperative ground resonance problem are (i) numerical integration of the equations of motion, and (ii) a smearing technique which involves a redistribution of damping over all the blades after one damper is considered inoperative. The reasoning for the second

approach is as follows: If the rotor has N blades then the total damping available in the rotor is Nc_i where c_i is the damping on one blade. If one damper is removed, the total damping becomes $(N - 1)c_i$. Thus, using the smearing approach, each blade in the rotor would be treated as if it had a lag damper rate equal to $c_i(N - 1)/N$. It is thus seen that this second approach analyzes a system which is quite different from the actual situation. The motivation behind this approach is to be able to use the standard Coleman and Feingold analysis for an isotropic rotor. In the results which follow, each of these techniques will be applied to a specific configuration and the ground resonance problem with one blade damper inoperative will be discussed in detail.

RESULTS

In order to illustrate the implications of one blade damper inoperative on the ground resonance characteristics of a single rotor helicopter, a set of parameters was chosen. These parameters were chosen so as to be in the general range of interest for single rotor helicopters and were such that the system was stable with all dampers functioning up to a rotor speed of 400 rpm. The parameters used in obtaining the results which follow are shown in Table 1.

Results for the system described by the parameters of Table 1 with all blade dampers operational are indicated in Figure 2. These results were obtained using the standard Coleman and Feingold approach. This approach results in only two equations which describe the rotor degrees of freedom regardless of the number of blades. Thus there are only four modes which result from the eigenvalue analysis. As can be seen from the upper portion of the figure, where the real parts of the eigenvalues are plotted as a function of rotor speed, the system is stable over the entire rotor speed range. The labeling on the various modes is intended for identification purposes only and is not meant to imply anything with respect to the character of the modes. In the lower portion of the figure is plotted the frequencies of the various modes as a function of rotor speed. The horizontal dashed lines represent the uncoupled hub modes and the slanted dashed lines represent the uncoupled blade modes. As can be seen, at the lower and higher rotor speeds the modes are essentially uncoupled, whereas for the intermediate rotor speeds a considerable amount of coupling is apparent.

When one blade damper is removed the results shown in Figure 3 are obtained. These results were calculated using the Floquet transition matrix method and indicate an instability for rotor speeds between 210 and 305 rpm. The nature of this instability can be

determined from an examination of the frequency plot in Figure 3. At the lower rotor speeds the frequencies for the modes labeled 3 and 6 correspond to the uncoupled blade frequency (in the fixed system) for the blade which has no damper. As the rotor speed is increased the frequency plot for mode 3 indicates coupling with the other modes, and at the higher rotor speeds the mode 3 curve has deviated from the uncoupled curve. At the higher rotor speeds the mode 5 curve is nearer the uncoupled blade frequency curve. This behavior indicates that there is a significant amount of coupling between the unstable mode 3 and the mode labeled 2 which is predominantly a hub mode. The conclusion here is that the indicated instability has the same character as a classical ground resonance instability, and that the instability involves a considerable amount of blade motion.

Figure 4 presents the results of a numerical integration of the equations of motion for a rotor speed of 255 rpm. This rotor speed corresponds to the point of maximum instability in Figure 3. As can be seen from the time histories in Figure 4, blade 1 which has no lag damper is experiencing large excursions. It should also be noted from this figure that determination of whether the system is stable or unstable requires considerable judgment on the part of the analyst. The blade 1 trace appears to be stable, whereas it is impossible to make a definite conclusion relative to the hub traces. This serves to illustrate the fact that time history solutions are less desirable than eigenvalue methods for determining the dynamic stability of systems. The problem with the time history solutions is that one can never be sure that the equations have been integrated over a sufficiently long period for the initial conditions chosen. If the initial conditions are not chosen so as to excite the mode of instability, an extremely long period of integration may be necessary. A further drawback of the numerical integration method is that, in general, it requires much more computing time than does the eigenvalue approach.

Results obtained using the smearing approach are illustrated in Figure 5. Note that although the mode labeled 3 becomes lightly damped, the system remains stable throughout the rotor speed range considered. The smearing technique is thus not recommended for treating the one-blade-damper-inoperative situation since it leads to unconservative results.

Suppose, however, that a designer were using the smearing technique and obtained the results shown in Figure 5. The logical approach to making the system more stable would be to add additional blade lag damping since this parameter is known to be quite effective in eliminating the classical ground resonance. The effect would be that the smearing technique would then indicate a sufficient stability

margin since it is simply an application of the classical analysis. Note, on the other hand, what the correct results, as obtained from the Floquet analysis and shown in Figure 6, would indicate. As can be seen from this figure, the increase of blade lag damping has no effect on the region of instability when one blade damper is inoperative. This result is not too surprising since it was observed from the time history traces of Figure 4 that the blade with no damper is responding more or less independently of the other blades. These results further strengthen the conclusion that the smearing technique should not be used for examining the one-blade-damper-inoperative ground resonance problem.

From the results of Figure 6 it might appear that the Army had imposed an impossible requirement on the helicopter designer. As will be shown, however, it is possible to eliminate the instability through proper selection of the parameters available to the designer. Since it has been shown that increasing the lag damping has no effect on the region of instability with one blade damper inoperative, the only other blade parameter which the designer can vary is the lag spring rate. Coleman and Feingold (1) have shown that, in the absence of lag damping, ground resonance is impossible if the blade lag frequency is greater than the rotor speed. This requirement is extremely conservative, however, if blade lag dampers are present.

Figure 7 illustrates the effect of lag spring rate on the region of instability when one blade damper is inoperative. The assumption was made here that the lag spring and lag damper were independent so that failure of the lag damper did not result in simultaneous failure of the lag spring. Thus the results of Figure 7 were obtained for a rotor in which each of the blades had the same spring rate, but one blade damper was inoperative. As can be seen, if the lag spring rate is made high enough the region of instability can be eliminated. The spring rate required is, on the other hand, much lower than the spring rate necessary to make the blade lag frequency equal to the rotor speed. It is felt, however, that the lag spring rate required is unrealistic and, further, the higher the spring rate the larger will be the vibratory loads transferred to the fuselage. Thus the lag spring does not appear to be the optimum parameter to use in eliminating the instability.

The other parameters available to the designer are the hub parameters. The hub stiffness and damping in the lateral and longitudinal directions may be changed by altering the landing-gear geometry and oleo characteristics. Further, the oleo stiffness and damping are usually adjustable in service so that these are potentially powerful parameters for the designer. Figure 8 shows the effect of lateral hub

HAMMOND

damping on the region of instability. This figure illustrates that hub damping can be used effectively in eliminating the instability. This can be attributed to the fact that the only means for the blade with no damper to experience any damping is through the hub coupling terms in the equations of motion. Thus the addition of hub damping tends to stabilize any unstable hub motions while providing damping for the large blade excursions.

The effect of variation of the lateral hub spring rate is shown in Figure 9. These results indicate that, if the stiffness of the hub is reduced, the instability can be eliminated. Again the indication is that the coupling between the rotor and hub is the key to eliminating the instability. Here the lower stiffness hub is allowed to respond more than the higher stiffness hub and thus a greater amount of hub damping can be transferred to the blade through the coupling terms.

As a final result, it was thought to be of interest to determine how much damping could be removed from one blade before encountering an unstable region. The results of these calculations are shown in Figure 10. From Table 1 it may be noted that each blade originally had a blade damper whose rate was 3000 ft-lb-sec/rad, and from the figure it is seen that an unstable region appears when the damping on one blade is reduced to approximately 1000 ft-lb-sec/rad. Thus, for this particular example, approximately two-thirds of the damping may be removed from one blade before an instability results.

CONCLUSIONS

Three main conclusions may be drawn from the results presented. First, the analyst must be careful to examine new user requirements to ascertain that the assumptions of existing analytical tools are not violated. If these assumptions are violated, modifications to the analysis must be made or new analyses must be formulated to handle the new requirements. Modifying the physical problem so that it fits existing analytical methods can lead to erroneous conclusions as evidenced by the smearing technique results.

The most effective means of eliminating the mechanical instabilities which occur with one blade damper inoperative appears to be through appropriate adjustment of the hub effective stiffness and damping characteristics. The approach seems to be to either reduce the hub stiffness to allow more hub response or increase the hub damping. A combination of reduced hub stiffness and increased hub damping will probably provide the most nearly optimum solution.

HAMMOND

Finally, the Floquet transition matrix method is an effective analytical tool for dealing with the one-blade-damper-inoperative ground resonance problem. The method provides the stability boundaries directly and thus eliminates the uncertainties associated with time history solutions.

REFERENCES

1. Coleman, R. P., and Feingold, A. M.: Theory of Self-Excited Mechanical Oscillations of Helicopter Rotors With Hinged Blades. NACA Report 1351, 1958.
2. Brooks, G. W.: The Mechanical Instability and Forced Response of Rotors on Multiple-Degree-of-Freedom Supports. Doctoral Dissertation, Princeton University, 1961.
3. Bielawa, R. L.: An Experimental and Analytical Study of the Mechanical Instability of Rotors on Multiple-Degree-of-Freedom Supports. Report No. 612, Department of Aeronautical Engineering, Princeton University, June 1962.
4. Hammond, C. E.: An Application of Floquet Theory to Prediction of Mechanical Instability. Paper presented at the AHS/NASA Ames Specialists' Meeting on Rotorcraft Dynamics, Moffett Field, California, February 13-15, 1974.
5. Peters, D. A., and Hohenemser, K. H.: Application of the Floquet Transition Matrix to Problems of Lifting Rotor Stability. Journal of the American Helicopter Society, Vol. 16, No. 2, April 1971, pp. 25-33.
6. Hohenemser, K. H., and Yin, S. K.: Some Applications of the Method of Multiblade Coordinates. Journal of the American Helicopter Society, Vol. 17, No. 3, July 1972, pp. 3-12.
7. Brockett, R. W.: Finite Dimensional Linear Systems. John Wiley and Sons, 1970.

HAMMOND

TABLE 1. PARAMETERS USED IN THE SAMPLE CALCULATIONS

Number of blades	4
Blade mass, m_b	6.5 slugs (94.9 kg)
Blade mass moment, S_b	65.0 slug-ft (289.1 kg-m)
Blade mass moment of inertia, I_b	800.0 slug-ft ² (1084.7 kg-m ²)
Lag hinge offset, e	1.0 ft (0.3048 m)
Lag spring, k_l	0.0 ft-lb/rad (0.0 m-N/rad)
Lag damper, c_l	3000.0 ft-lb-sec/rad (4067.5 m-N-s/rad)
Hub mass, m_x	550.0 slugs (8026.6 kg)
Hub mass, m_y	225.0 slugs (3283.6 kg)
Hub spring, k_x	85000.0 lb/ft (1240481.8 N/m)
Hub spring, k_y	85000.0 lb/ft (1240481.8 N/m)
Hub damper, c_x	3500.0 lb-sec/ft (51078.7 N-s/m)
Hub damper, c_y	1750.0 lb-sec/ft (25539.3 N-s/m)

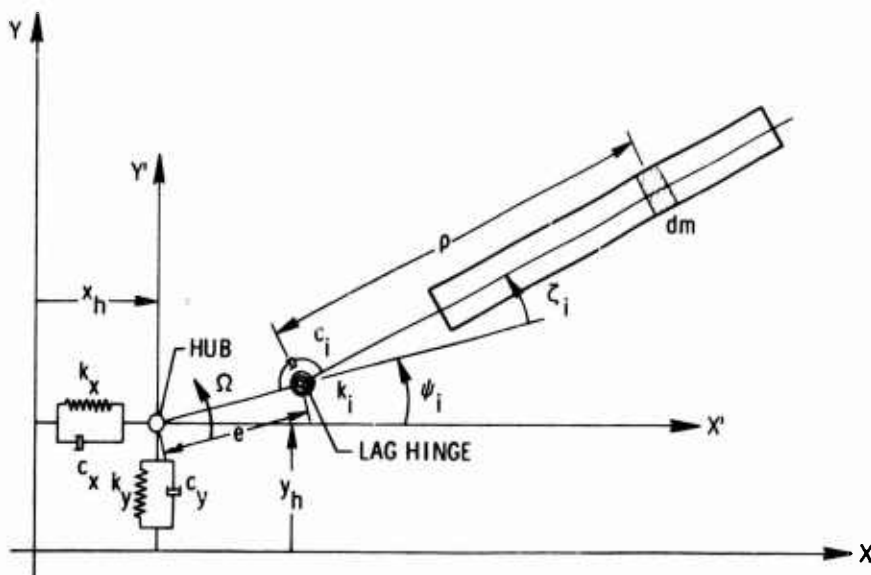


Figure 1. Mathematical representation of the rotor and hub.

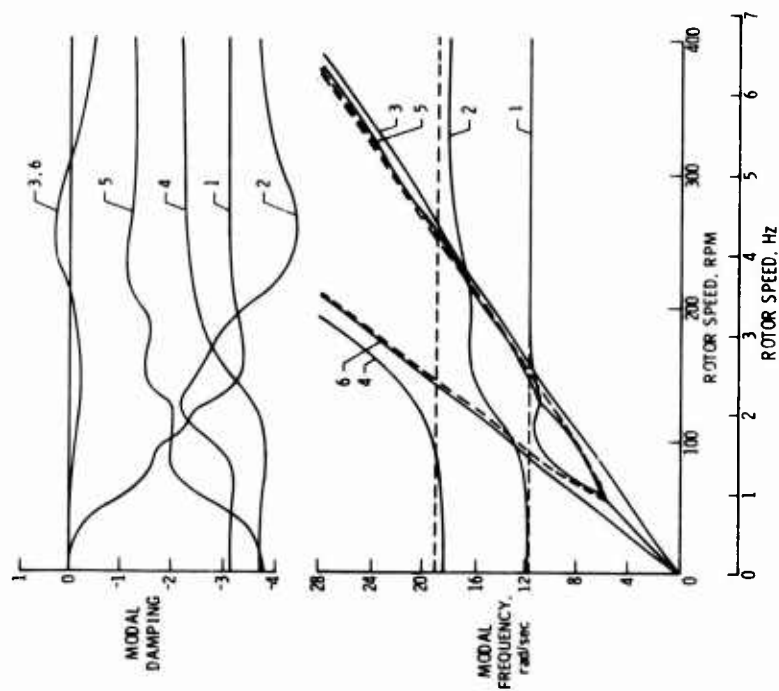


Figure 3. Modal damping and frequencies for nonisotropic hub, one blade damper inoperative. Frequencies plotted in the fixed system.

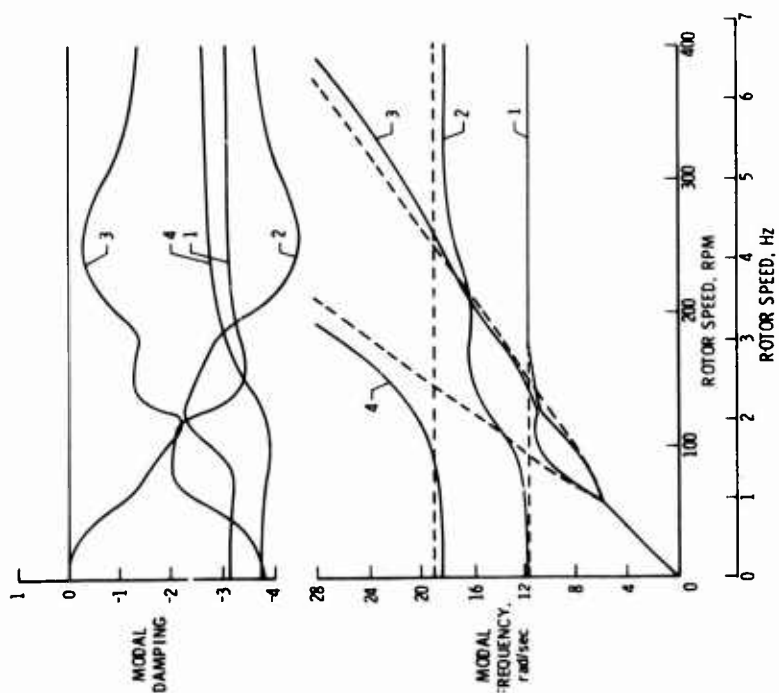


Figure 2. Modal damping and frequencies for nonisotropic hub, all blade dampers working. Frequencies plotted in the fixed system.

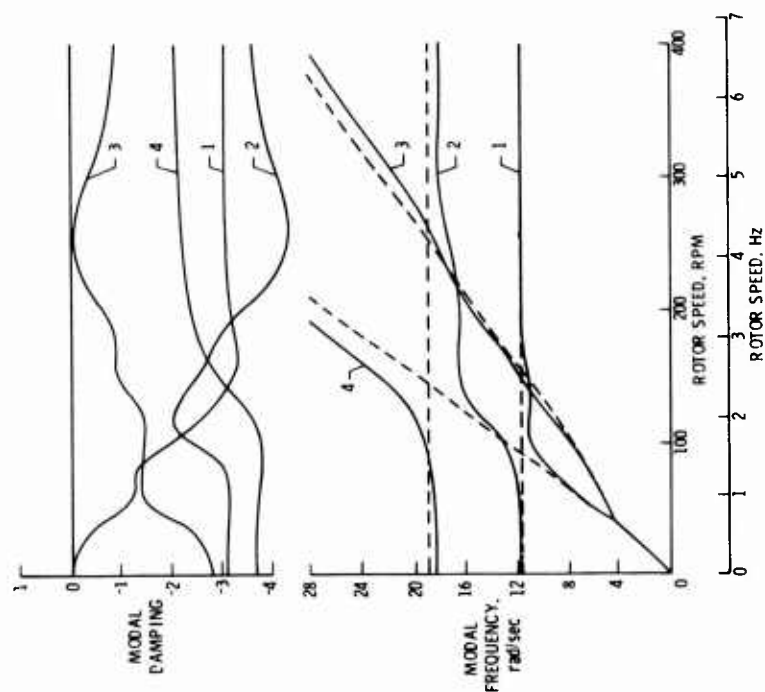


Figure 5. Modal damping and frequencies obtained for nonisotropic hub, one blade damper inoperative, using the smearing technique.

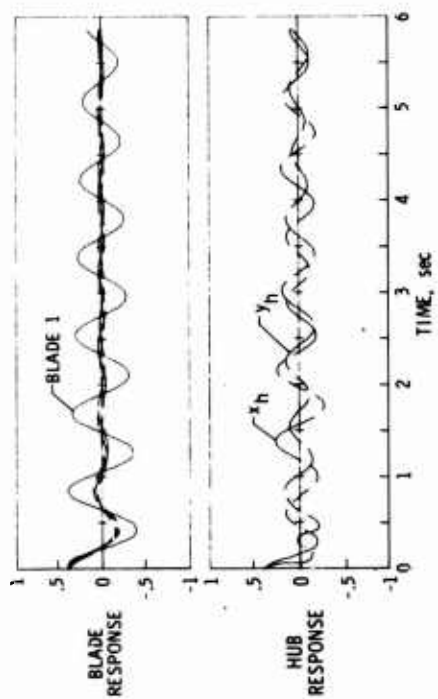


Figure 4. Time history calculations for nonisotropic hub, one blade damper inoperative, $\Omega = 255$ rpm.

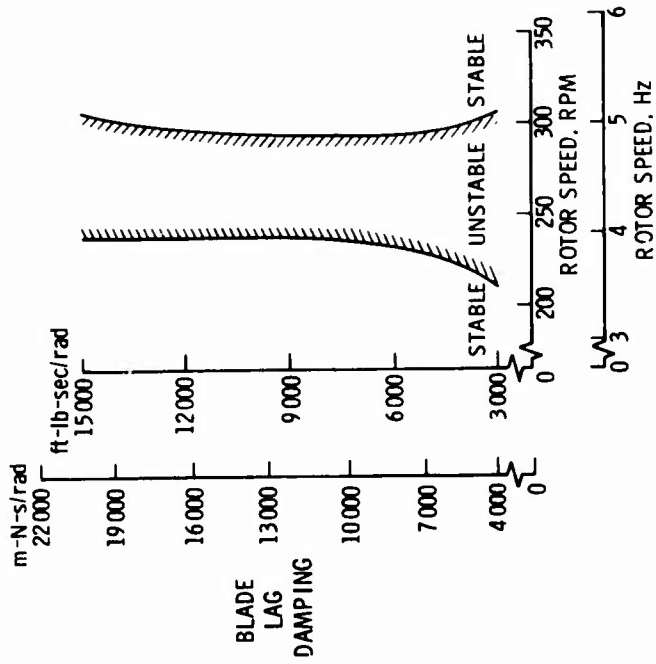


Figure 6. Instability region as a function of blade lag damping for the non-isotropic hub and one blade damper inoperative.

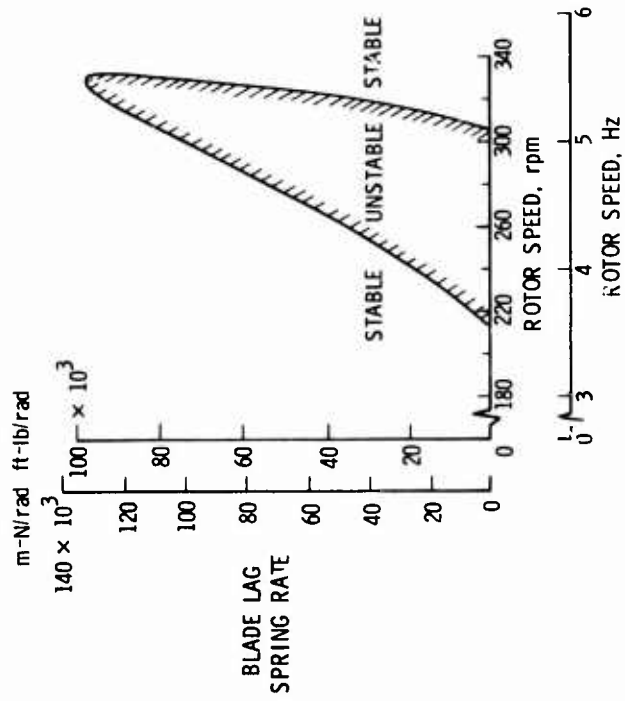


Figure 7. Effect of blade log spring rate on region of instability for one blade damper inoperative.

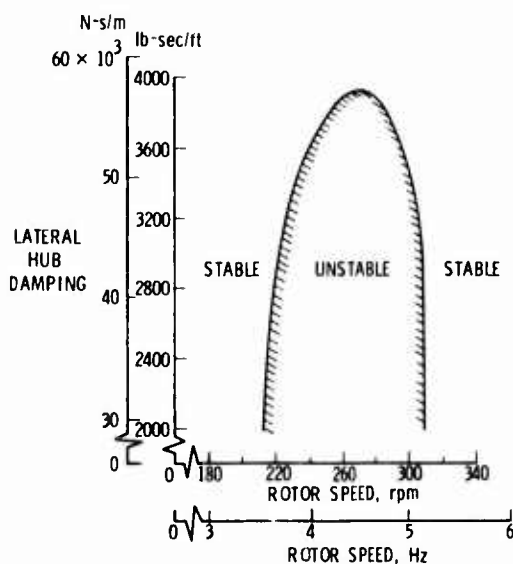


Figure 8. Effect of lateral hub damping on region of instability with one blade damper inoperative.

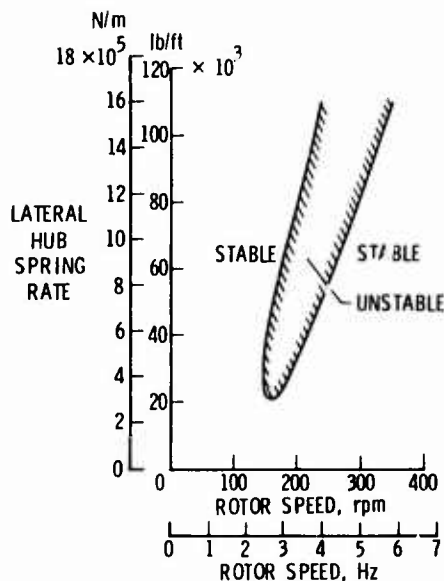


Figure 9. Effect of lateral hub spring rate on region of instability with one blade damper inoperative.

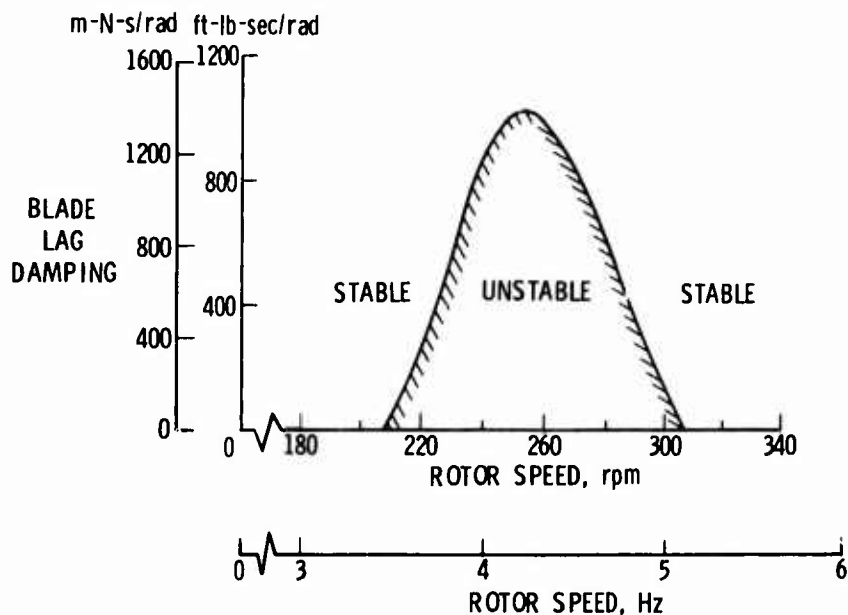


Figure 10. Region of instability resulting from reduced lag damping on one blade.

DETECTION AND IDENTIFICATION OF TRACE QUANTITIES OF
ORGANIC VAPORS IN THE ATMOSPHERE BY ION CLUSTER MASS
SPECTROMETRY AND THE IONIZATION DETECTOR SYSTEM

CHARLES S. HARDEN, Ph.D.
THOMAS C. IMESON, Ph.D.
CHEMICAL LABORATORY
EDGEWOOD ARSENAL, ABERDEEN PROVING GROUND, MARYLAND

I. INTRODUCTION

The study of recently discovered proton ion/dipole clusters in the gas phase (1-4) has been extended in connection with the search for new concepts in chemical agent (and other atmospheric contaminants) detection. The result has been a new analytical technique called Ion Cluster Mass Spectrometry (ICMS) (5). ICMS is akin to the technique of chemical ionization (CI) mass spectrometry in that both concepts rely upon the mass analysis of ions that are products of ion-molecule reactions occurring in a reactant gas in the ion source. ICMS employs moist air as the reactant gas at an ion source pressure of 10 torr to 1 atmosphere. The sample may be introduced into a clean, moist air stream or the sample plus reactant gas may constitute ambient atmosphere. Sample/reactant gas ratios may be as low as 10 ppb.

Mechanisms, rate constants, and enthalpies of formation for ion clusters $H^+(H_2O)_n$ (derived from ion molecule reactions initiated by the primary ions N_2^+ and O_2^+) have been established where the primary ionization source has been electron irradiation (6), in flowing afterglow experiments (7), and in corona discharges (5,8,9). For the ICMS technique, the hydrated proton species serve as precursors for clusters $H^+(H_2O)_n(A)_m$ which represent the qualitative and quantitative indicators of the trace compound A. The mechanism of formation of $H^+(H_2O)_n(A)_m$ is not dependent upon the method of formation of primary ions, i.e. EI or electrical discharges. The sensitivity

Preceding page blank

of the technique is, however, dependent upon the rate of formation of the primary ions, the water vapor concentration, and the pressure of the reactant gas. A corona discharge was selected for this work because it is an extremely intense source of primary ions.

Some of the results of the ICMS studies have been applied to the establishment of the response mechanisms in the Ionization Detector System (IDS) now under exploratory development for the Army and the Air Force. The basic field unit, as it exists to date, has been developed by Minneapolis Honeywell, Inc. by empirical methods. The mechanism by which the IDS selectively responds to chemical agents in the atmosphere has not been fully understood. The construction details, sensitivity, and specificity of the system have been given in contractor reports (10).

II. EXPERIMENTAL

Figure 1 is a diagram of the ICMS apparatus. Ions produced in a corona discharge or as a result of ion-molecule reactions (formed at room temperature, $\sim 25^{\circ}\text{C}$) diffuse through a pinhole into a vacuum chamber where they are focused into a quadrupole mass spectrometer. Ion detection is accomplished with an electron multiplier coupled to a high speed picoammeter or electrometer amplifier.

Air or nitrogen containing trace quantities of water vapor and the substance to be analyzed flow into the ion source - a corona discharge cylinder. The discharge is established between a 90% Pt - 10% Rh wire, .08 mm diameter, and a stainless steel cylinder, 12 mm I.D., located coaxial to one another. The discharge current is maintained at 10 μa (for all source pressures) with a 6-kv variable power supply placed in series with a 4.7 megohm current-limiting resistor. Ions exit the source chamber by diffusion through an orifice, 35 microns in diameter, located in the center of S and are subsequently mass analyzed.

Flow rate through the ion source is held constant for all pressures. Source pressures are varied to allow a control of the collision rate and the subsequent appearance and/or disappearance of ionic species of interest. Analytical data is obtained when the source pressure is held constant and the sample concentration is variable.

A sketch of a detection cell in the IDS in the configuration used in this work is shown in figure 2. The system operates at atmospheric pressure. Air, heated to 60°C (the normal operating

temperature of the IDS), is pumped through the cell at any desired flow rate. The cell is placed in a small oven so that 60°C is maintained throughout the cell. The air stream passes over a 1-Ci β -source (titanium tritide coated foil). Primary ions (N_2^+ , O_2^+ , and O_2^-) are formed near the β -source by electron impact or electron attachment. A sequence of ion-molecule reactions follows and equilibrium among ionic clusters is rapidly established. The mixture of air and ions then flows through a series of baffles to a Faraday cup ion collector. One end of the Faraday cup is a grid to allow the air to flow through. The β -source which is in electrical contact with the center manifold stud can be biased either positively, negatively, or maintained at zero potential with respect to the collector. The ion currents are measured with a picoammeter.

Trace concentrations of samples are prepared in a continuous flow system by using air dilution of vapor from the compound of interest.

III. RESULTS AND DISCUSSION

A. Ion Cluster Mass Spectrometry

Specific ionic species observed in an ion cluster mass spectrum of dimethyl methylphosphonate are given in the table below.

Identification of DMMP Spectral Elements

Ion	m/e	I/I (%)
$H^+(H_2O)_3$	55	2.5
$H^+(H_2O)_4$	73	8.0
$H^+(H_2O)_5$	91	7.0
$H^+(H_2O)_6$	109	1.5
$H^+(DMMP)$	125	11.1
$H^+(H_2O)_7$	127	1.0
$H^+(DMMP)(H_2O)$	143	14.1
$H^+(DMMP)(H_2O)_2$	161	8.5
$H^+(DMMP)_2$	249	43.2
Other	-	3.1

This spectrum was taken using air as the reactant gas at a source pressure of 140 torr, [DMMP] = .6 ppm, and $[H_2O] \sim 100$ ppm. As can

be seen in the table, indicators of the trace compound appear at m/e of $[MW + 1 + n(18)]$, e.g., $H^+(DMMP)(H_2O)_n$, and at m/e of $[n(MW) + 1]$, e.g., $H^+(DMMP)_n$, where n is 1 or 2.

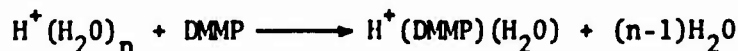
Ions that appear under the category "other" in the table constitute background and can be attributed to clusters formed with methanol and acetone - solvents used in cleaning the apparatus. Consequently, no fragment ions need or can be assigned to the spectral elements.

Figures 3 and 4 illustrate the normalized ion intensities versus pressure of DMMP and dimethyl sulfoxide (DMSO) systems in the pertinent high pressure regions. Concentrations of DMMP and DMSO are approximately the same (.6 and .7 ppm respectively) and $[H_2O] \sim 100$ ppm. It will be noted that indicators of DMMP appear at lower pressures than those for DMSO and that the rates of formation of $H^+(DMSO)_2$ and $H^+(DMSO)$ are about the same. Without evidence to the contrary, the mechanism of formation of ion clusters for both DMMP and DMSO is assumed to be the same. However, for ion clusters formed from DMMP, the formation of the proton-dimer has predominance over the formation of the proton-monomer, while for DMSO these two reactions have a parity in their probability of occurrence.

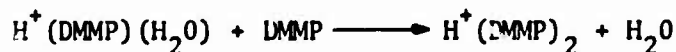
Dipole moments for DMMP (11) and DMSO (12) are respectively 3.62 and 3.96 Debye units; whereas their polarizabilities, computed from refractive indices, are 10.6 \AA^3 and 7.9 \AA^3 . An explanation for the apparent greater stability and ease of formation of clusters containing DMMP compared to DMSO may be interpreted as a greater influence of polarizability over dipole moment. However, the refractive indices and densities of the two substances are such that the difference in their polarizabilities is primarily a function of their molecular weights. Consequently, while dipole moment and polarizability certainly influence cluster formation, it is assumed that the greater number of vibrational degrees of freedom resident in the DMMP molecule impart a larger stability to its clusters when compared to those of DMSO. This would seem to indicate, that within certain unspecified limits, the larger a dipolar molecule the more stable and readily formed will be its ion clusters.

The quantification of the ICMS technique and its sensitivity with respect to trace atmospheric impurities is determined by studying ion cluster formation as a function of the impurity concentration

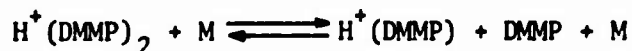
at constant pressure. Figure 5 illustrates the dependence of various ion cluster concentrations upon [DMMP] at 140 torr. The plotted points indicate experimental data and the smooth curves represent the results of numerically fitting kinetic equations to the data. For the curve fitting procedure, several reaction mechanisms were considered. The mechanism resulting in the best fit is slightly different from one previously proposed (5) but, nevertheless, is consistent with the qualitative predictions that can be drawn from figures 3 and 4. This mechanism and calculated rate constants are given below.



$$k = 1.85 \times 10^{-8} \text{ cm}^3/\text{sec}$$



$$k = 4.06 \times 10^{-8} \text{ cm}^3/\text{sec}$$



$$k_f = 2.77 \times 10^{-13} \text{ cm}^3/\text{sec}$$

$$k_r = 7.4 \times 10^{-25} \text{ cm}^6/\text{sec}$$

The average deviation of the calculated points from experimental data is 12.4%.

At concentrations near the threshold sensitivity the indicators of the trace impurities are $\text{H}^+(\text{A})(\text{H}_2\text{O})$, where A is an impurity molecule. At higher concentrations, a few ppm, the ion cluster mass spectrum consists only of ion clusters $\text{H}^+(\text{A})_n$, where n is usually 1 and 2.

The sensitivity of the ICMS technique to organic trace atmospheric impurities is, at this time, about .01 ppm. The sensitivity is limited by the ion transmission capability and signal-to-noise characteristics of the mass spectrometer presently in use.

B. Ionization Detector System

It is now well understood that the charge carriers in ionized air, at or near atmospheric pressure, are in the form of ion clusters,

regardless of the source of primary ionization. It is, therefore, evident that responses in the IDS are caused by changes in densities of ionic clusters as a function of trace impurity concentration as air flows through the detection cell (figure 2). The current changes usually appear as increases (increase in positive ion cluster concentration) when an impurity is introduced into a clean moist air stream.

Both positive and negative primary ions are formed in the β -source region of the detector cell and react with neutral molecules to form positively and negatively charged clusters. In the absence of atmospheric impurities the positive clusters are $H^+(H_2O)_n$ and the negative clusters are primarily $O_2^-(H_2O)_n$ (13). Under these conditions, more positive clusters than negative clusters reach the collector per unit time. The net result is a positive current measured at the collector.

When impurities are present in the air reactions between the impurity molecules and the hydrate clusters result in the replacement of water in the clusters by the impurity. The clusters thus produced are larger in size and, presumably, more stable.

1. Qualitative Response Studies

At the beginning of a series of parametric studies on the IDS, qualitative sniff tests were run for approximately 60 compounds. In general, the IDS responses were positive, i.e., there were increases in the net positive cluster concentrations at the collector. Some compounds showed an increase in the net negative ion current (examples are most chloro compounds) and some showed no change in cell current.

Examination of the responses of an homologous series of primary alcohols (methyl through n-butyl) showed that the larger the impurity molecule, the larger the cell response. It was also noted that large concentrations of ammonia caused no cell response. Ammonia is known to react very well to form ammonia-water ion clusters (9) and the cluster size distribution is essentially the same as that with only water present. The conclusions drawn from these observations are a) the larger molecules form larger clusters b) since larger clusters have small diffusion velocities normal to the flow streamlines, they have a greater probability of surviving the flow region and increasing the collector current.

2. Quantitative Response Studies

The conclusions derived from the qualitative observations were tested by measuring the response of the IDS to quantitative mixtures of several homologous series of impurity compounds.

Primary alcohols (methanol through n-pentanol) were studied at concentrations of 25 and 50 ppm. At each concentration the order of response was n-pentanol>n-butanol>n-propanol>ethanol>methanol. (An ICMS analysis of methanol, ethanol, and n-propanol, each at a concentration of 25 ppm in air, shows the alcohols readily react to form ionic clusters $H^+(ROH)_n(H_2O)_m$. The distribution of n shifts toward larger values, i.e., the average mass of the clusters increases with increasing molecular weight.)

A series of CW agent simulants (phosphonate esters) was studied as functions of concentration and flow rate. The series consisted of diisopropyl methylphosphonate (MW 180), diethyl ethylphosphonate (MW 166), dimethyl methylphosphonate (MW 124), and dimethyl hydrogenphosphonate (MW 110). The results for DIMP and DMMP are shown in figure 6. The other members of the series were left out for the sake of clarity. As can be seen, the response increases to saturation with concentration. At any given concentration and flow rate the response due to DIMP>DEEP>DMMP>DMHP.

Three amines were tested quantitatively and the same behavior was noted; namely, the response of diisopropyl amine>diethyl amine>>ammonia.

The IDS response shows different response characteristics to acids. At low concentrations the response decreases (increases negatively) markedly. As concentration is increased further the response passes through a minimum and then increases toward saturation. The negative ion clusters are formed faster at lower concentrations and tend toward saturation sooner than the positive ions. The magnitude of responses (both positive and negative) due to the acids studied is butyric acid>acetic acid.

All of the quantitative work done so far shows that, within a homologous series, the larger molecules result in larger ion clusters and produce greater IDS responses. Thus, the quantitative studies seem to verify that the size of the ionic clusters, therefore, diffusion, plays a predominant role in the IDS response mechanism.

3. Diffusion Studies

The effect of cluster size can be demonstrated by studying the IDS response to water vapor at various air flow rates. The positive, I^+ , negative, I^- , ion currents (+ and - 200 volts source bias) were measured as a function of flow rate, Q , up to about 8.5 lit/min. The relationship between I and Q was empirically found to be

$$I = K \cdot Q^{3/2}$$

In the vicinity of the source of ionization the concentrations of positive and negative ions is dependent upon temperature, pressure, and β -source activity. These parameters are held constant so that the ion density in the ion source region is constant. The ion loss mechanism in this region of the detector is primarily recombination which reduces both positive and negative ion concentrations by the same amount.

To a first approximation the cell geometry is considered as a straight cylindrical tube of volume V (the effective internal volume of the baffle section of the cell) and radius r (half the average spacing of the various components of the baffle section). A certain number of ions per unit area, N_0 , is introduced into the baffle section. Only radial diffusion is considered since axial diffusion does not result in a loss of ions. The one dimensional diffusion equation with the approximation for small distance is

$$n(x,t) = \frac{N_0}{2\sqrt{\pi Dt}} \exp\left[-\frac{x^2}{4Dt}\right] \simeq \frac{N_0}{2\sqrt{\pi Dt}}$$

An ion is considered to be lost and therefore not able to reach the ion collector when it has diffused through the distance, r (the radius of the approximating tube). The ion concentration at the collector end of the tube is then given by

$$n(t) = \frac{N_0}{2\sqrt{\pi Dt}}$$

Integration of the above along the radial distance of diffusion provides the number of ions per unit area at the collector.

$$\int_0^r n(t) dx = N(t) = \int_0^r \frac{N_0 dx}{2\sqrt{\pi Dt}} = \frac{N_0 r}{2\sqrt{\pi Dt}}$$

The equality of ratios for ions per unit area and ion concentration can be made

$$\frac{N(t)}{N_0} = \frac{n(t)}{n_0} = \frac{r}{2\sqrt{\pi D t}}$$

Therefore

$$n(t) = \frac{n_0 r}{2\sqrt{\pi D t}}$$

The ion current at the collector is

$$I = n(t) q Q$$

q is the electronic charge (1.6×10^{-19} coul).
Therefore,

$$I = \frac{n_0 r q Q}{2\sqrt{\pi D t}}$$

The time, t , is the residence time of particles in the tube and is V/Q . Substituting for t , the positive or negative ion current at the collector is

$$I = \frac{n_0 r q Q^{3/2}}{2\sqrt{\pi D V}}$$

Thus, the average ionic diffusivity can be determined from the slope, K of a plot of I vs $Q^{3/2}$. In a typical IDS cell, $r \approx .038$ cm and $V \approx .19$ cm³. Dimensions of the ion source region are such that recombination of positive and negative ions is an efficient mechanism for ion loss when ion concentrations are greater than 1.5×10^7 ion/cm³; below this concentration recombination is unimportant. Therefore, the value chosen for n_0 is 1.5×10^7 ions/cm³ for both positive and negative ions. Then,

$$D = \left[\frac{n_0 r q}{2K} \right]^2 \cdot \frac{1}{\pi V} = \frac{1.61 \cdot 10^{-23}}{K^2}$$

when I is measured in amperes and Q in liters/min.

Figure 7 represents the results of diffusion studies (I vs $Q^{3/2}$) for moist air. At low humidity, separate and distinct curves are obtained for positive and negative ions. This indicates that the negative ions are smaller than the positive ions. When the humidity is increased, ionic diffusivity decreases (larger ions are formed), and the curves coincide; oppositely charged ions are essentially the same size. The values of diffusivities calculated from these data agree very well, considering the complexity of the actual cell geometry, with typical ion diffusivities ($.06 \text{ cm}^2/\text{sec}$) reported in the literature (14) and are summarized below.

Rel. Humidity (25°C)	$D^+ (\text{cm}^2/\text{sec})$	$D^- (\text{cm}^2/\text{sec})$
15%	.075	.094
63%	.052	.052
92%	.030	.030

The observed decrease in D with humidity is consistent with the well known fact that the average size of hydrated ion clusters increases with $[\text{H}_2\text{O}]$.

The conclusions drawn from these studies is that differential diffusion of ionic clusters from an air stream predominates in the physics of the response mechanism. A recent, independent study of the IDS mechanism (15) has substantiated the conclusions of this work with respect to the diffusion mechanism.

IV. SUMMARY

A. Ion Cluster Mass Spectrometry

Irrespective of the nuances associated with ion cluster stability, the fact remains that the technique producing them offers an opportunity for application to the detection and analysis of trace compounds in the gas phase. The ability to use atmospheric gases as reactants considerably extends the practical applicability of this technique. Its incorporation with gas chromatographic analysis of trace compounds (where N_2 is a common carrier gas) has many possibilities. Several features that enhance its general attractiveness are: 1) the absence of a spectrum complicated with fragment ions, 2) high mass peaks directly related to the

molecular weight of the compounds, and 3) sensitivity well below the ppm concentration range.

The ultimate sensitivity of ICMS is yet to be determined. As mentioned previously, the sensitivity has been limited by the sensitivity of the mass spectrometer. Another mass spectrometer has recently been incorporated into the apparatus. The new system should increase sensitivity by at least an order of magnitude. The mass range, one to about 1200 amu, provides the capability of observing large clusters of large organic molecules.

B. Ionization Detector System

It has been shown that the charge carriers in the IDS are ionic clusters of water and trace gas molecules. The ability to use such a system as a detector of trace atmospheric impurities is obtained from the facts that 1) ion clusters are readily formed in moist air, 2) the presence of trace impurities tends to rapidly form larger, more stable clusters, 3) the smaller diffusivities of the large ionic clusters provide a means of separating the clusters according to size.

Although the IDS responds (with varying sensitivity) to many compounds it is essentially non-responsive to "normal" atmospheric pollutants (e.g., NO_x , SO_2 , H_2S , etc.) and other compounds of low molecular weight. Therefore, interferences from these substances and others likely to be found in a realistic environment of its intended use are minimal. Whatever inadequacies such a system suffers in its lack of specificity for detection are made up in its inherent sensitivity and its simplicity.

ACKNOWLEDGEMENT

The authors acknowledge the technical efforts of J. A. Parsons and D. J. Buckley who gathered most of the experimental data. We are especially grateful to Mrs. Linda Ambrose for her clerical support in preparing the manuscript.

REFERENCES

1. Tickner, A. W. and Knewstub, P. F., Mass Spectrometric Studies of Ions in Glow Discharges. IV. Water Vapour. J. Chem. Phys., 38, 464 (1963).
2. Kebarle, P. and Godbole, E. W., Mass Spectrometric Study of Gases at Near-Atmospheric Pressures. J. Chem. Phys., 39, 1131 (1963).
3. Narcisi, R. S. and Bailey, A. D., Mass Spectrometric Measurements of Positive Ions at Altitudes from 64 to 112 Kilometers. J. Geophys. Res., 70, 3787 (1965).
4. Kebarle, P., High-Order Reaction - Ion Clusters and Ion Solvation (with references). Ion Molecule Reactions, Edited by J. L. Franklin, Vol. 1, Chap. 7, Plenum Press, New York, 1972.
5. Imeson, T. C. and Harden, C. S., Chemical Ionization Ion Cluster Mass Spectrometry. Edgewood Arsenal Technical Report, EATR 4642, Edgewood Arsenal, MD, May 1972.
6. Good, A., Durden, D. A., and Kebarle, P., Ion Molecule Reactions in Pure Nitrogen and Nitrogen Containing Traces of Water at Total Pressures of 0.5-4 Torr. Kinetics of Clustering Reactions Forming $H^+(H_2O)_n$. J. Chem. Phys., 52, 212 (1970), and Mechanism and Rate Constants of Ion-Molecule Reactions Leading to Formation of $H^+(H_2O)_n$ in Moist Oxygen and Air. Ibid., p. 222.
7. Ferguson, E. E. and Fehsenfeld, F. C., Water Vapor Ion Cluster Formation in the D Region. J. Geophys. Res., 74, 5743 (1969).
8. Shahin, M. M., Use of Corona Discharges for the Study of Ion-Molecule Reactions. J. Chem. Phys., 47, 4392 (1967).
9. Harden, C. S., A Mechanism for the Formation of Electrically Charged Ammonia-Water Clusters in the Condensation Nuclei Personnel Detector. Edgewood Arsenal Technical Report, EATR 4569, Edgewood Arsenal, MD, November 1971.
10. Anderson, W. E., Schluter, B. C., and Harris, W. A., Multi-Agent

HARDEN and IMESON

Detector System (U). CONFIDENTIAL Report, Contract AFATL-TR-72-212 Final Report, Honeywell, Inc., Minneapolis, Minnesota, 1972.

11. Kosolapoff, G. M., Dipole Moments of the Lower Dialkyl Alkyl-phosphonates, J. Chem. Soc. (London), 1954, 3222 (1954).
12. Nelson, R. D., Lide, D. R., and Maryott, A. A., Selected Values of Electric Dipole Moments for Molecules in the Gas Phase. NSARDS-NBS 10, US Government Printing Office, Washington, D. C., 1967.
13. Burke, R. R. and Miller, W. J., Research Studies on the Condensation Nuclei (CN) Personnel Detector. Third Quarterly Progress Report, Contract No. DAAA15-68-C-0469, AeroChem Research Laboratories, Inc., Princeton, NJ, April 1969.
14. McDaniel, E. W. and Mason, E. A., The Mobility and Diffusion of Ions in Gases. John Wiley and Son, 1973.
15. McClure, B. T. and Waletzko, J. A., Ionization Detector System (IDS) Mechanism. Unpublished work. Honeywell, Inc., Minneapolis, Minnesota.

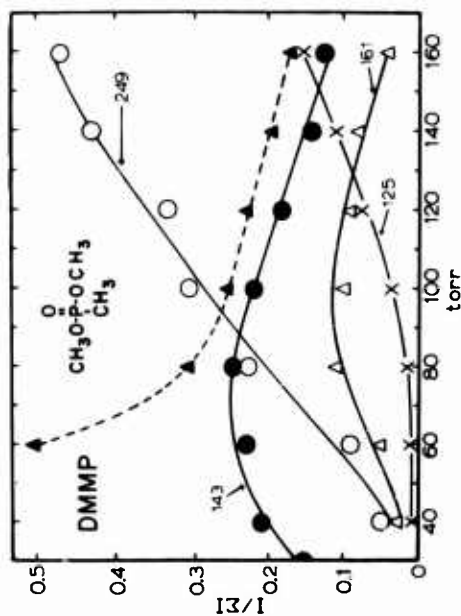


Figure 3: Fractional ion intensities in DMMP vs pressure.

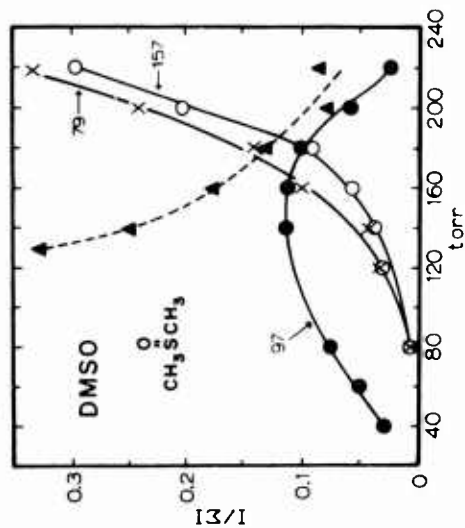


Figure 4: Fractional ion intensities in DMSO vs pressure.

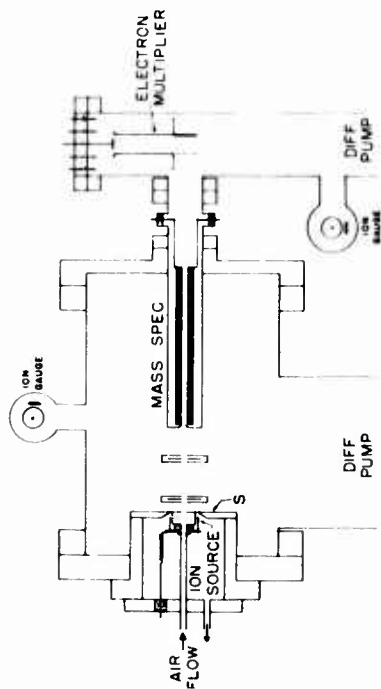


Figure 1: Schematic diagram of the ICMS apparatus.

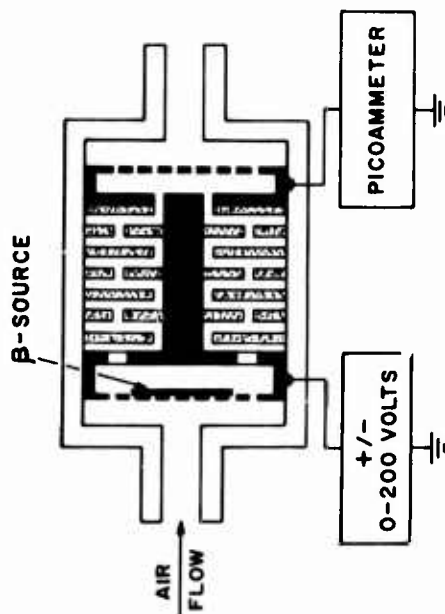


Figure 2: Schematic diagram of an IDS cell.

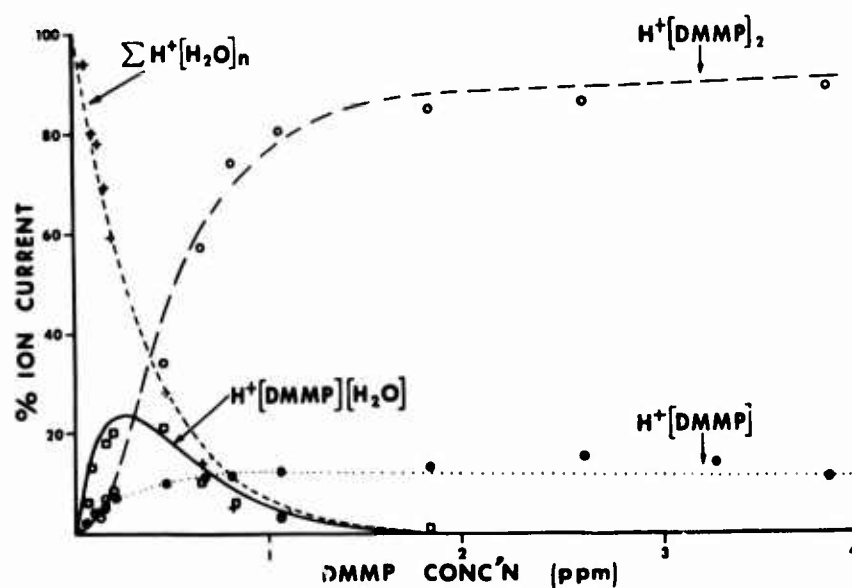


Figure 5: ICMS relative ion intensities vs [DMMP]

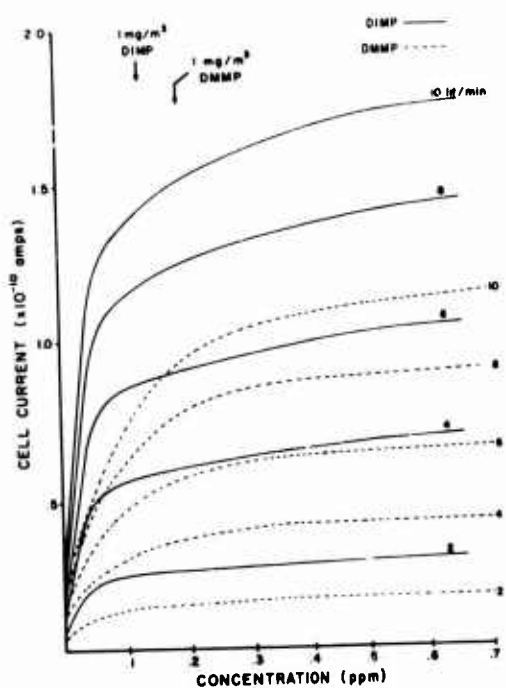


Figure 6: IDS response vs conc'n at several flows

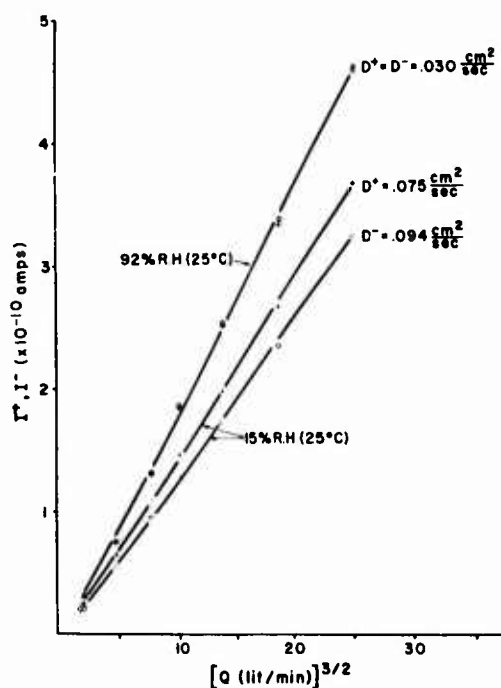


Figure 7: + & - IDS response as a function of flow

CELL AND ORGAN CULTURE SYSTEMS TO EVALUATE BIODEGRADATION
AND BIOCOMPATIBILITY OF POLYMERS

DR. ANDREW F. HEGYELI and
JAMES C. EATON, Jr.
US ARMY MEDICAL BIOENGINEERING R&D LAB
FORT DETRICK, FREDERICK, MARYLAND 21701

INTRODUCTION

For more than 30 years a large number of organic polymers have been used as surgical repair materials. The specifications for these materials vary with their intended use. For instance, for fabrication of heart valves or artificial hearts, the materials need to be chemically and mechanically stable in the biological environment for long periods of time. In other cases, such as hemostatic agents, tissue adhesives for wound closure, surgical suture materials, and certain types of bone replacements, it is desirable to employ biodegradable materials. This latter type of implant polymer is expected to maintain its mechanical properties for limited periods of time and is programmed to degrade, yielding non-toxic metabolites which are eliminated through normal excretory routes as the implant is gradually replaced by the body's own tissues. For the past several years, efforts in the U.S. Army Medical Bioengineering Research and Development Laboratory have been directed toward synthesis and evaluation of tissue adhesives, biocompatible polymers, and biodegradable surgical repair materials. Whether the aim is to develop a biostable or biodegradable material, it is important that the degree of biodegradation be accurately evaluated and that the degree of toxicity of the degradation products be assessed prior to implantation in patients.

In practice, no polymeric material has been found to date which is perfectly compatible with the tissues of the recipient. The methods to determine the relative compatibility of the polymers consist of short and long term implantation of the polymers in animals and of histological examination of the tissue reaction.

Preceding page blank

Chemical degradation for polymer characterization is estimated routinely by in vitro chemical tests and biodegradation by long term in vivo implantation tests in animals. While yielding important information, these tests suffer from the following drawbacks:

1. In vitro chemical tests, consisting of the determination of the degradation of the polymers in saline at different pH values and at elevated temperatures, or in mixtures of organic solvents and saline according to US Pharmacopeia specifications for plastic containers, may not correlate with the in vivo performance of the polymer¹.

2. In vivo implantation studies in animals require several months or even years, and the amount of biodegradation is difficult to quantify for slowly degrading materials unless large numbers of animals are used. On the other hand, long term implantation of radio-labeled material is very costly and requires special housing and safety precautions, especially since a certain amount of the degradation product might be excreted through the exhaled air.

To circumvent these difficulties, three quantitative in vitro techniques were developed in our laboratory. These methods are based on the degradation of radio-labeled polymers in (1) chick embryo organ homogenates; (2) organ cultures; (3) primary cell cultures.

Cell culture techniques are also used for biocompatibility studies, the end point being changes of the cell morphology and cell death in the vicinity of the polymer.

Cell and organ culture systems can also be used to determine the biocompatibility or toxicity of polymers.

Histological and electron microscopic examination of the exposed cultures and monitoring of the influence of the biodegradation products on living cells by time lapse cinemicrography give good indications of the biocompatibility or toxicity of the test materials.

IN VITRO BIODEGRADATION TESTS

Three systems were developed to assess the biodegradation of polymers.

1. Chicken embryo organ homogenates
2. Chicken embryo liver organ cultures
3. Chicken embryo lung primary cultures

Fourteen day old Leghorn chicken embryos are used in all three systems. The selection of chicken embryo tissues for our studies was made on the basis of their high metabolic rate and their unlimited availability as a homogeneous, easily controlled experimental system with respect to age of the animals at the time of experiment. The embryos are sacrificed in a laminar flow hood by decapitation using aseptic technique, rinsed in Eagle's basic salt solution (BSSE) containing 125 units of penicillin-streptomycin solution per ml, and stored in ice until used for the different in vitro systems.

(1) Chicken Embryo Organ Homogenates

Materials and Methods

Organs such as lungs, liver, skeletal muscle are separately homogenized in a cell homogenizer and suspended in minimum essential medium of Eagle, supplemented with 10% chicken serum and 5% chicken embryo extract. The medium contains L-glutamine and appropriate amounts of antibiotics. In a routine test, 10 mg of ^{14}C -labeled polymer powder, prepared in a standardized way^{2,3} is introduced into the homogenate and shaken for varying periods of time in an incubator at 39°C in order to expose the polymer to the action of enzymes and other biological substances present in the organ.

Control samples are incubated in tissue culture medium or in heat denatured homogenates.

At the end of the reaction the samples are centrifuged, filtered from the residual polymer, and aliquots of the filtrate are counted in a scintillation spectrometer. The radioactivity of the filtrate is compared with that of the original polymer and expressed as percent degradation.

Results

Homologous series of (alpha-2-cyanoacrylates)- ^{14}C were prepared by our laboratory² for use as biodegradable tissue adhesives. The biodegradation rates for the different cyanoacrylate derivatives were determined in a 24 hour experiment in embryo liver homogenate, and gave the following results:

Table I
Degradation of Alpha-2 Cyanoacrylates in Liver Chicken
Embryo Homogenates After 24-Hours Incubation at 37°C

2-CYANOACRYLATE DERIVATES	PERCENT DEGRADATION
METHYL	37.9
ETHYL	1.65
PROPYL	0.78
ISO-BUTYL	0.67
ISO-AMYL	0.54
N-BUTYL	0.36
HEPTYL	0.13

These data show the relative rates of degradation for a homologous series of poly(alkyl 2-cyanoacrylate) powders and indicate that the methyl derivative degrades most rapidly and that the rate of degradation decreases with increased length of the alkyl chain. These in vitro data are in agreement with results from rat and dog long term in vivo implantation studies which showed that methyl cyanoacrylate degrades more rapidly than the butyl derivative⁴. The degradation of poly(lactic acid)-¹⁴C polymers³ was also studied in this system. The results of this test are summarized on Table II.

Table II
Degradation of Poly(lactic acid) Derivatives in Chicken Embryo
Liver Homogenates in 24-Hours at 37°C

POLY(LACTIC ACID) DERIVATIVES	PERCENT DEGRADATION
L(+) PLA	0.27
DL-PLA	0.51

The data indicate that poly[L(+) lactic acid] is less susceptible to degradation than the dl polymer. This is expected because of the crystalline nature of the poly[L(+) lactic acid] compared to the dl polymer. These results are also in accord with previously reported results obtained by an in vivo method^{5,6,7} and by gross observations of poly(lactic acid) implants.

The organ homogenate system is also used for the preparation of soluble degradation products of polymers. The tissue culture media containing these metabolites and the solubilized materials such as small molecular weight polymers, plasticizers, fillers, etc. are utilized for the biocompatibility test of these polymer extracts in the time lapse cinemicrography study as described under "Dynamic Biocompatibility Test".

(2) Chicken Embryo Liver Organ Cultures

The development of techniques for evaluating biodegradation in short term organ cultures represents a further improvement and extension of the tissue homogenate in vitro studies. These organ culture techniques are even more rapid and sensitive than the embryo organ homogenate systems. Furthermore, they make it possible not only to determine the quantity but also the quality of biodegradation. Using these systems, it is possible to make a number of specific determinations to explore the precise nature of the biodegradation process and the toxicity and fate of the biodegradation products.

Materials and Methods

The organ culture biodegradation test is performed in the following manner: Using sterile technique, the organs of 8 to 14 day old Leghorn chicken embryos are removed in a laminar flow hood immediately following decapitation of the embryos. The liver is cleared from adhering tissues and cut into fragments approximately 0.5 x 0.5 cm in size. These fragments are then rinsed with basic salt solution of Eagle (BSSE) containing 125 units of penicillin streptomycin solution per ml, and placed in roller tubes with Trowell T8 tissue culture medium, supplemented with 10 percent chick serum, 5 percent chicken embryo extract, 125 units of penicillin streptomycin mixture and 2 mM of L-glutamine per ml. The ^{14}C -labeled candidate polymers [β - ^{14}C on the cyanoacrylates, and on the poly(lactic acid)] are added to the cultures in powder form, and the cultures are then incubated in CO_2 at 37°C with rotation at a rate of 1 RPM for varying periods of time. The test tube caps are kept loose to enable gas exchange. The gas mixture in the incubator is 5 percent CO_2 and 95 percent air in all cases. At the end of the test the cultures are centrifuged, the supernatant fluid filtered and the samples are counted in a liquid scintillation spectrometer, and the radioactivity of the filtrate is compared with that of the original polymer and expressed as percent degradation.

Results

As can be seen in Table III, organ cultures give even higher rates of degradation than when the polymer is exposed to an equal amount of tissue homogenate for the same period of time. Thus the organ culture test is even more sensitive and rapid and is particularly valuable for the evaluation of slowly degradable polymers.

Table III
Percent Degradation of Cyanoacrylate Derivatives in Chicken Embryo Liver Homogenate and Embryo Liver Organ Culture in 24-Hours at 37°C

CYANOACRYLATES	CELL HOMOGENATE	ORGAN CULTURE
METHYL	37.9	52.2
ETHYL	1.65	3.1
PROPYL	0.78	1.99
ISO-BUTYL	0.67	1.52
N-BUTYL	0.36	0.73

In vivo studies²⁷ in rats also indicated that the radioactively labeled metabolites were not stored in the tissues but gradually excreted by normal excretory routes. At 90 days following implantation, 0.73 percent of the label was found in the feces, 3.73 percent in the urine, 83.5 percent was residual, and 12.04 percent was unaccounted for. It was assumed that the radioactive label unaccounted for in these studies might possibly have been excreted as CO₂. These data were obtained in long term implantation studies requiring the use of special metabolic cages and a large number of animals.

In the case of exploratory biocompatibility studies, when large numbers of polymer derivatives need to be evaluated rapidly, it was desirable to collect the above data in a more rapid fashion. We therefore explored the feasibility of using organ cultures to collect similar types of information on the metabolic fate of polymers. The filtered degradation products of different cyanoacrylate derivatives obtained in organ homogenates were introduced into liver, organ cultures and as can be seen in Table IV, the label was taken up by the liver cultures from the medium at varying rates.

Table IV
Percent Uptake of the ^{14}C Label by Chicken
Embryo Liver Organ Culture in 24-Hours at 37°C

CYANOACRYLATES	PERCENT UPTAKE
METHYL	7.7
ETHYL	5.9
ISO-BUTYL	16.4
N-BUTYL	24.1

Experiments were also carried out to determine whether the liver is able to completely metabolize the soluble cyanoacrylate and poly(lactic acid) derivatives to CO_2 . For this study the labeled polymers were incubated in a closed system as shown in Fig. 1.

The organ cultures and the labeled polymers were incubated in a Warburg reaction vessel, and the volatile metabolites trapped in Hyamine, or in $\text{Ba}(\text{OH})_2$ solution to collect only the $^{14}\text{CO}_2$. The results of these studies are summarized in Table V and show that a certain amount of the label is found in the gaseous phase both in the case of cyanoacrylates and poly(lactic acids).

Table V
Percent Recovery of the ^{14}C Label from Poly(methyl-cyanoacrylate
and Poly(lactic acid) after 24-hours Incubation
in Chicken Embryo Liver Cultures

POLYMER	MEDIUM + LIVER	AS CO_2
POLY(METHYL-2-CYANOACRYLATE)	99.8	0.21
POLY(DL-LACTIC ACID)	98.8	1.13

(3) Chicken Embryo Lung Primary Cultures

Materials and Methods

Chicken embryo lungs are rinsed with BSSA and cut with scalpel into small squares about 1 x 1 mm. These explants are then cultured

on chicken plasma clot in Falcon plastic Petri dishes for 3 days in Medium-199 supplemented with 10% chicken serum, 5% embryo extract and antibiotics. The labeled polymers are introduced into the system in powder form as before, but in the case of slowly degrading materials the incubation is extended to 72 hours or even longer as required. The filtrate of the reaction mixture is then freeze-dried in small polyethylene bags to prevent loss of the material and the whole sample is oxidized and the $^{14}\text{CO}_2$ is trapped in the scintillation cocktail.

Results

This method enabled us to obtain statistically significant data from very slowly degrading polymers as well. Some of the data are summarized in Table VI.

Table VI
Biodegradation of Methyl-Cyanoacrylate and Poly(lactic acid)
Derivatives in Chicken Embryo Primary Lung
Cultures After 72-Hours Incubation at 30°C.

POLYMER	NO. SAMPLES	% DEGRADATION	STD. DEV.
MeCNA	2	43.2	3.3
DL PLA	6	4.74	2.41
L(+) PLA	2	3.02	1.14

Using methyl-cyanoacrylate as a positive control, it is possible to compare results from different experiments. In this case the degradation rate of this internal standard is corrected to 50% and the values of the experimental samples are interpolated accordingly. It has to be mentioned that the methyl cyanoacrylate gave only small variations in values in 72 hour parallel experiments. The corrected data of these experiments are shown in Table VII.

Table VII
Comparative Biodegradation Rates of Poly(lactic acid)
Derivatives, Polyester Materials and Butyl(Cyanoacrylate)
in Chicken Embryo Lung Organ Culture in 72-Hours at 37°C
with Methyl Cyanoacrylate as Internal Standard

POLYMER	% DEGRADATION	STD. DEV.
MeCNA	50.0	8.0
DL PLA	10.1	8.7
L(+) PLA	3.5	1.3
PCL	0.44	0.16
PPL	3.82	0.04
BuCNA	0.70	0.04

IN VITRO BIOCOMPATIBILITY TESTS

Biocompatibility can be assessed by in vivo long term implantation studies in different animals species and by in vitro cytotoxicity systems. The end point in the implantation experiments is determined by the evaluation of the cellular response to the implant in terms of the number of neutrophils, plasma cells, giant cells and the thickness of the fibrous capsule around the implant^{8,9}. Several in vitro tests have also been suggested, based on different tissue culture systems. Normal and malignant cell lines, diploid cells and primary cell cultures are used for this purpose by a number of investigators. The end points of the evaluation are changes in cellular morphological or cell viability as determined by vital stain uptake.

(1) Histologic and Electron Microscopic Studies

The introduction of cell culture systems for implant biodegradation studies in our laboratory, and development of the organ culture and primary cell culture methods for this purpose offered a possibility for screening exposed cultures for signs of cellular damage.

Histological evaluation of the organ culture fragments indicated that highly toxic degradation products induce obvious cellular changes, leading to pyknotic nuclei, cytoplasmic vacuolization and cell death. On the other hand, borderline toxic events can easily be overlooked in this evaluation. Electron microscopic studies of the

preparation, on the other hand, indicated that even in the case of low cellular toxicity, the cell damage can be easily assessed.

Fig. 2 is an electron micrograph of a control preparation showing normal subcellular structure and diffusely arranged nuclear chromatin structure.

Fig. 3 is a parallel embryo liver culture following short exposure to the degradation products of a methyl-cyanoacrylate polymer implant. Histological examination of this specimen did not indicate definite evidence of toxicity, but the electron microscope revealed a fragmented chromatin pattern, denatured nucleoplasm, lack of glycogen granules, and mitochondrial damage. The lysosomes; lipofuscin; denuded, rough endoplasmic reticulum, and free ribosomes indicate extensive, non-specific cellular damage.

The combination of organ homogenate methods and primary cell culture studies enabled us to develop a dynamic tissue toxicity test.

(2) Dynamic Biocompatibility Test

The organ homogenate system is capable of providing biodegradation products expected to be formed in the body following implantation. The filtrate of the reaction mixture is then tested in primary cell cultures, and the effect of the degradation products is monitored by time lapse cinemicrography.

Materials and Methods

Chicken embryo lung explants are cultured on round glass cover slips placed in plastic Petri dishes for 3 days in CO₂ incubator. The medium is the same as described for the primary cell cultures. The explants are placed on chicken plasma clot to accelerate the outgrowth of cells. At three days the cells do not reach confluency, and both the fibroblastic and the epithelial-like cells are clearly distinguishable. The cover slips are then transferred into a Dvorak-Stotler controlled environmental culture chamber and the chamber is attached to the perfusion unit. This consists of a 250 ml i.v. infusion bottle containing the tissue culture medium and connected to the chamber through an intravenous injection set provided with air inlet bacterial filter, drip chamber and drip tube and plastic tubing with rubber connector for supplementary injection of the polymeric degradation products obtained from the organ homogenate system. The chamber is filled and is incubated in an air stream incubator on the

microscope stage plate. The perfusion rate is 3-5 ml/hour and the cells are photographed through phase contrast or Nomarski interference contrast objective on 16 mm film at 2 Fr/minute time-lapse.

Results

In this system marginal toxic events can be overcome by the cells, but toxic metabolites bring about serious cell damage and cell death. Since embryonic primary cultures are considered to be the most sensitive to toxic substances¹⁰ even low toxicity can be detected. The change in the mobility of fibroblastic cells, the ballooning of the cells, the vacuolization and finally the total cessation of cell motion is shown in the short sequence to be demonstrated at the Army Research Conference. Motility as a crucial indication of life is a very sensitive indicator, and thus gives a sharper end point than changes in the morphology of the cells.

Fig. 4 is a scanning electron micrograph of a critical point dried primary culture exposed for 1/2 hour to methyl cyanoacrylate, and Fig. 5 is a parallel culture exposed to methyl cyanoacrylate for 2 hours. The changes in the gross morphology of the cells are not prominent, but the damage to the cell surface as a result of low toxicity metabolite exposure is clearly visible.

Discussion

Polymers have long been used as surgical implants. Hundreds of thousands of people are benefiting from them. Systematic tests of polymers for biocompatibility started in 1949, when Aries¹⁰ implanted nylon sutures in dogs and observed the tissue reaction to the polymer and the effect of the bio-environment on the plastic. In 1945 Blum^{11,12} introduced tissue culture systems for the in vitro screening of cytotoxicity of polymers. Since then, many methods have been developed to test for biocompatibility on the basis of in vivo implantation^{1,8,12-18} and in vitro cell cultures.^{12,19-23} The first in vitro biodegradation test was developed in our laboratory.²⁴⁻²⁶ The three tests described in this paper are routinely used for the preliminary screening of surgical polymers developed by our laboratory or for candidate materials obtained commercially. The data obtained with these methods are in good correlation with in vivo implantation studies and thus provide an important tool in materials development.

HEGYELI & EATON

The authors wish to express their appreciation to Dr. R.K. Kulkarni, Dr. Clarence W.R. Wade, Mr. J.G. Dillon and Mr. G. Brandes for the preparation of PLA, cyanoacrylate and polyester materials used in these studies, and to Mr. M.A. Gonda of Litton Bionetics, Inc., Frederick Cancer Research Center for the assistance in the preparation of the scanning electron micrographs.

REFERENCES

1. Jaeger, R.J. and Rubin, R.J. Science, 170:460, 1970.
2. Leonard, F., Kulkarni, R.K., Brandes, G., Nelson, J. and Cameron, J.J. J. Appl. Polym. Sci., 10:259, 1966.
3. Kulkarni, R.K., Pani, K.C., Neuman, C., and Leonard, F. Arch. Surg., 93:839, 1966.
4. Pani, R.K., Gladieux, G., Brandes, G., Kulkarni, R.K., and Leonard, F. Surgery, 63:481, 1968.
5. Kulkarni, R.K., Moore, E.G., Hegyeli, A.F., and Leonard, F. J. Biomed. Mat. Res., 5:169, 1971.
6. Brady, J.M., Cutright, D.E., Miller, R.A., Battistone, G.C. J. Biomed. Mater. Res. 7:155, 1973.
7. Cutright, D.E., Hunsuck, E.E. Oral Surgery, 31:134, 1971.
8. Sewell, W.R., Wiland, J., Craver, B.N. Surg. Gynecol. Obstet., 100:483, 1955.
9. Postlethwait, R.W., Arch. Surg. 101:489, 1970.
10. Aries, L.J. Surgery, 9:51, 1941.
11. Blum, G. Proc. R. Soc. Med., 38:169, 1945.
12. Blum, G. Br. J. Surg., 33:245, 1945.
13. Salthouse, T.N. and Williams, J.A. J. Surg. Res., 9:481, 1969.
14. Salthouse, T.N. and Willigan, D.A. J. Biomed. Mat. Res., 6:105, 1972.
15. Homsy, C.A., J. Biomed. Mat. Res., 4:341, 1970.
16. Guess, W.L. and Rosenbluth, S.A. J. Pharm. Sci., 54:1545, 1965.
17. Ocumpaugh, D.E. and Lee, H.L. J. Macromol. Sci. Chem. A 4(3):595, 1970.
18. Oppenheimer, B.S., Oppenheimer, E.T. and Stout, B.T. Proc. Soc. Exp. Biol. 79:366, 1952.
19. Rosenbluth, S.A., Weddington, G.R., Guess, W.L. and Autian, J. J. Pharm. Sci., 54:156, 1965.
20. Johnsson, R.I. and Hegyeli, A.F. Ann. N.Y. Acad. Sci. 146:66, 1968.
21. Hegyeli, A.F., Johnsson-Hegyeli, R.I.E. Nature, 213:4171, 1967.
22. Johnsson-Hegyeli, R.I.E. and Hegyeli, A.F. Trans. Amer. Artif. Int. Org., 14:48, 1968.
23. Johnsson-Hegyeli, R.I.E. and Hegyeli, A.F. J. Biomed. Mater. Res. 3:115, 1969.
24. Hegyeli, A.F. Fed. Proc., 29:744 Abs., 1970.
25. Hegyeli, A.F. In Vitro, 7:267, 1972.
26. Hegyeli, A.F. J. Biomed. Mater. Res., 7:205, 1973.
27. Cameron, J.L., Woodward, S.C., Pulaski, E.J., Sleeman, H.K., Brandes, G., Kulkarni, R.K., and Leonard, F. Surgery, 58:424, 1965.

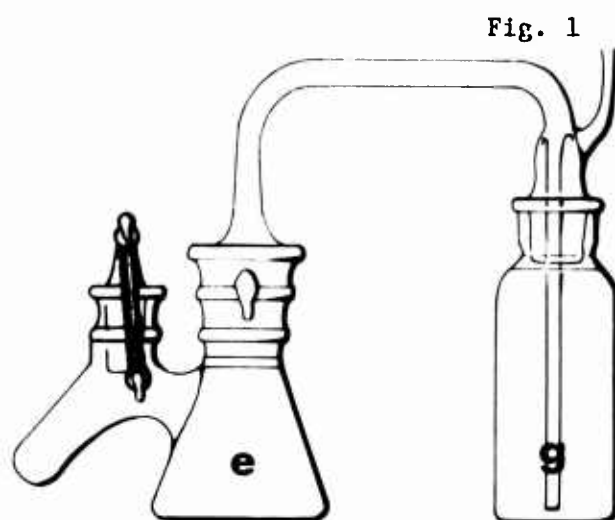


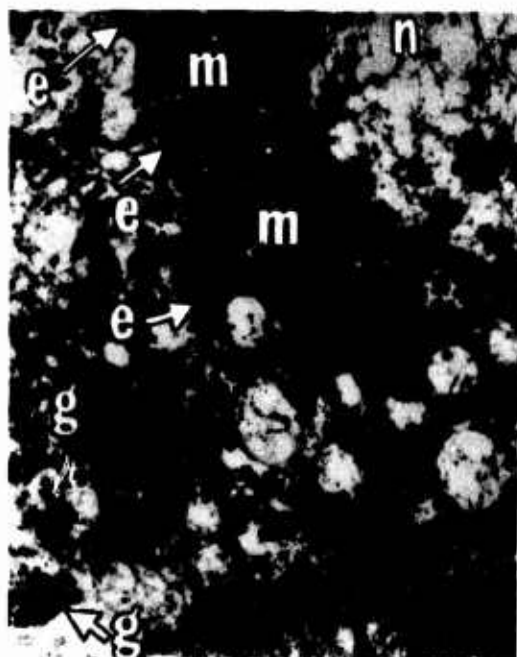
Fig. 1

Closed Organ Culture System, Used in the Study of Volatile Metabolites

e = Explant
g = CO₂ Trap

Fig. 2

Fig. 3



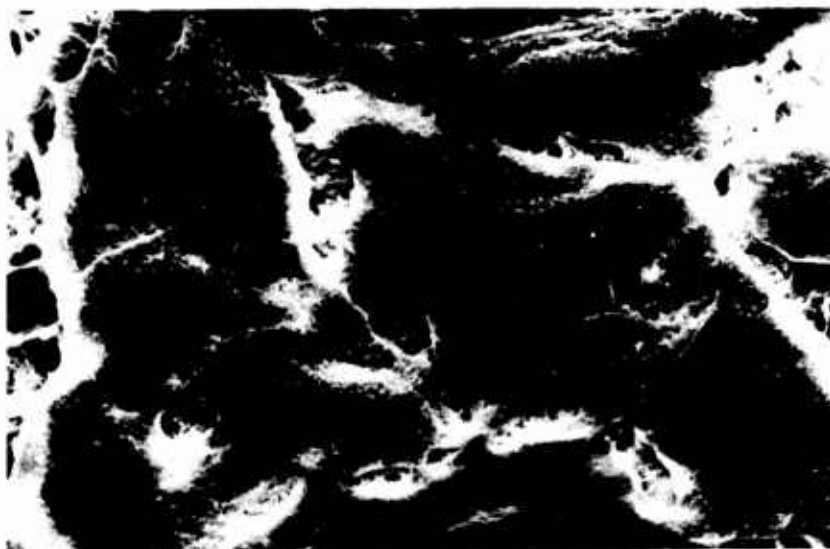
Electron Micrograph of a Control Liver Slice. 13,000 X



Parallel Liver Slice Exposed to Methyl-cyanoacrylate for 8 hours. 4,900 X

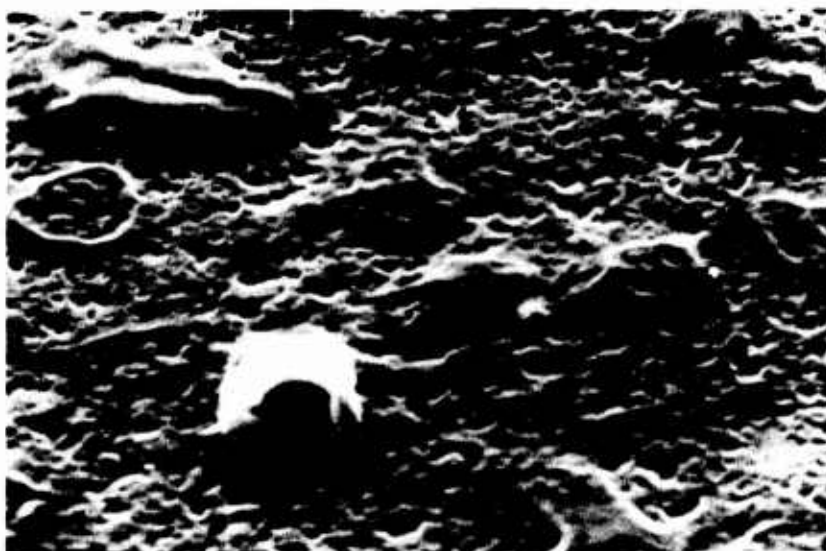
n = Nucleus m = Mitochondrion e = Endoplasmic Reticulum
g = Glycogen Granule v = Vacuole pm = Plasma Membrane

Fig. 4.



Chick Embryo Primary Culture Exposed
For 1/2 Hour to Methyl-cyanoacrylate 2500 X

Fig. 5.



Surface of Flattened Fibroblast Like Cell
After 2H Exposure to Methyl-cyanoacrylate 7000 X

A SEA ICE TERRAIN AND MOBILITY MODEL

WILLIAM D. HIBLER III, PhD
USA COLD REGIONS RESEARCH AND ENGINEERING LABORATORY
HANOVER, NEW HAMPSHIRE

With the advent of a new generation of air cushion vehicles, often called surface effects vehicles (SEV), the nature of the Arctic pack ice as a vehicle terrain has taken on military and commercial significance. Because of this significance, studies have recently been carried out at USA CRREL in order to effectively characterize sea ice terrain for purposes of SEV mobility, as well as for other types of vehicles. These studies have involved aerial stereo photography, airborne laser profilometry, and ground measurements performed on the pack ice surface.

It is clear from such studies as well as from previous research that the most salient characteristics of pack ice terrain are pressure ridges which form due to alternating dilation and compression of the ice cover. These pressure ridges often reach several meters in height and kilometers in length, and occur with frequencies up to 10 to 15 per kilometer. Pressure ridges typically form in regions of thin ice (see for example Weeks et al. (1), Kovacs, (2)) caused by the refreezing of leads, which form when the pack ice cracks and pulls apart. To convey some feeling for these obstacles we show in Figure 1 a ground view of a first year pressure ridge (in the Beaufort Sea) and in Figure 2, an aerial view of sea ice which illustrates the frequency and intersecting nature of pressure ridges.

In addition to pressure ridges on the pack ice surface there often exist rubble fields, consisting of relatively small ground up ice chunks, rafted ice consisting of the over and under thrusting of two ice sheets, and large leads or openings in the ice

Preceding page blank

HIBLER

which impede land based vehicles but which are easily crossed by amphibious vehicles.

In order to parametrize this terrain, studies at USACRREL have involved terrain characterization both by spectral densities (Hibler and LeSchack (3), Hibler (4)) of the surface roughness and by the statistical properties of discrete pressure ridges (Hibler et al. (5) Mock et al. (6)). For trafficability studies discrete characterization of pressure ridges has been found to be most useful, whereas for ride characteristics the spectral densities have been found to be more useful since they provide a description of smaller scale roughness such as the undulating nature of multi-year ice due to alternating melt ponds and low rounded hummocks. In particular the parametrization of ridging characteristics has been reduced to a one parameter model which gives good estimates of the number of ridges per kilometer at any height level. This model together with extensive laser profile data has made possible the characterization of regional and temporal variations of ridging in the western Arctic Basin (Hibler et al. (7)).

Using such information on ridge characteristics together with additional information on ridge extents it is possible to develop a trafficability model for vehicles traveling the pack ice and to determine regional variations in trafficability. Such a development is carried out in this paper. To this end, the paper is divided into two parts: 1) A review of the one parameter ridge model and the regional variation of ridging in the western Arctic Basin, as derived from laser profile data; 2) The development of a trafficability model for Arctic sea ice terrain.

ONE PARAMETER RIDGE DISTRIBUTION MODEL

To characterize ridging we first need to be able to predict the number of ridges at any height level encountered per unit length along a straight line path. For this purpose, the following ridge height probability density function derived by Hibler et al. (5) has been found to agree well with experimental data

$$P(H) dH = 2 \mu_h \sqrt{\frac{\lambda}{\pi}} \frac{e^{-\lambda H^2}}{\operatorname{erfc}(h\sqrt{\lambda})} dH \quad (1)$$

HIBLER

where $P(H) dH$ is the number of ridges per kilometer with heights between H and $H+dH$, μ_h is the number of ridges per kilometer above height h , an arbitrary cut off height; λ is the distribution shape parameter which can be determined from $\langle H_h \rangle$, the mean ridge height of all ridges above h . Note that equation 1 represents a two parameter model, with the two most common parameters used being μ_h and $\langle H_h \rangle$.

With a relatively small loss of accuracy equation 1 may be converted to a one parameter model (Hibler et al. (7)) by assuming that it may be written in the form

$$P(H) dH = F(\gamma_h) e^{-\lambda(\gamma_h) H^2} dH \quad (2)$$

where F and λ are considered to be functions of a parameter γ_h (called ridging intensity) defined by $\gamma_h \equiv \mu_h \lambda$. It can be shown that γ_h scales approximately linearly with the volume of deformed ice per unit area (Hibler et al. 1974). Integrating equation 2 gives an implicit relation between F , λ and γ_h . For a second equation to determine the two unknowns F and λ given γ_h , we assume that $F(\gamma_h)$ is a linear function of $\sqrt{\gamma_h}$ with the constant parameters determined by fitting a regression line through F and γ_h values calculated from several sets of laser data using the two parameter model.

An illustration of how well this model fits laser profile data is given in Figure 3. In this figure we show the predicted and observed number of ridges above 4, 6 and 8 ft. The data consists of 16 sets of laser profile data, each containing approximately 40 km of track, taken at different geographical locations in the Arctic Basin in November 1970. To illustrate geographical variations in the ridging; we show in Figure 4, the location and ridging intensity of each of the 16 samples along with rough contours of the ridging intensity. Figure 4 in conjunction with Figure 3 may be used to estimate the number of ridges at different height levels for different regions in the Arctic Basin. More detailed studies of regional and temporal variation (Hibler et al. (7)) indicate that the geographical ridging variations in Figure 4 are quite typical for winter, although in a heavily ridged year the ridging intensity magnitudes may be up to 70% higher than shown in the figure.

TRAFFICABILITY MODEL

In order to develop a trafficability model some additional information, beside the frequency of ridges above different heights along a straight line path is needed. In particular, information on ridge orientation and extents is required. Consequently besides assuming the one parameter model describes the distribution of ridge heights we make the following additional assumptions:

1) The orientation of ridges is assumed to be random (Mock et al. (6)); that is for a given area there are approximately equal lengths of ridges oriented at any angle. As a corollary to this assumption of random orientation it can be shown that the ridge density R_D (total length of ridges per square kilometer) is related to the ridge frequency μ along a linear path by $R_D = (\pi/2) \mu$.

2) For the extent of ridges we assume that the distribution of ridge lengths is given by some distribution function $P_L(L)dL$, with first and second moments $\langle L \rangle$ and $\langle L^2 \rangle$. Examination of aerial photo mosaics has indicated that as a first approximation $P_L(L)dL$ is negative exponential in form with $\langle L \rangle$ of the order of 1 to 2 km.

Based on these characteristics, a typical sea ice obstacle model could be created in the following conceptual manner. First a "box" of random ridges is created using the height distribution $P(H)dH$ and the length distribution $P_L(L)dL$. Then, taking a two-dimensional region, a number of these random (random both in x and y coordinates and in orientation) ridges would be placed on the surface until the total length of ridges equaled the desired ridge density R_D .

Certain qualifications to the above model should be noted. One qualification is that a profile along a ridge rapidly varies in height. Consequently, although the use of the lateral extent distribution is reasonably valid for larger vehicles, small vehicles such as snowmobiles may be able to slip through a ridge at many locations because of small holes in the ridge. This is discussed in somewhat more detail by Hibler and Ackley (8).

Another point is the variation of $P_L(L)dL$ for different ridge heights. Since such variations may occur, trafficability analyses should generally provide for determining the sensitivity to changes in $P_L(L)dL$. The assumption of random orientation is also one that, although reasonable over large areas (tens of kilometers), breaks down over small areas (several kilometers) where

ridges often seem to occur in a linear pattern (Hibler (4), Mock et al. (6)). The height distribution, however, does appear to be valid for all ridging systems.

Having described the terrain to consist of a set of randomly oriented linear obstacles with a known height distribution, we can now carry out a simple trafficability calculation. To do so we will consider a completely maneuverable vehicle that can clear any ridge up to a height h . We will then calculate the ratio of the total distance traveled to the component of distance in a desired direction. This ratio D_T/D_{SL} turns out to be a unique function of the number of ridges per unit length higher than h , denoted by μ_h , and of the first and second moments ($\langle L \rangle$ and $\langle L^2 \rangle$) of the ridge extent distribution function. Consequently, using statistics compiled from laser profile data (as in Figs. 3 & 4) D_T/D_{SL} ratios be estimated for different regions of the Arctic.

First, an analytic calculation for the trafficability ratio D_T/D_{SL} neglecting ridge intersections is performed. This analytic calculation is similar to that discussed by Bekker (9) with extensions to include obstacles of different lengths. This calculation is useful for illustrating how the sampling biasing affects the results; i.e., long ridges perpendicular to the desired motion are encountered more frequently than short ones. Second, the basic calculation for D_T/D_{SL} is performed, using a Monte Carlo method which takes into account cul-de-sacs and ridge intersections. The point at which the results obtained with the Monte Carlo method approach those obtained with the analytic model is also important since it indicates at what point ridge intersection problems become negligible. In general, as might be expected, the results indicate that D_T/D_{SL} very rapidly increases as the ridge intersections begin to be encountered.

Analytic calculation of D_T/D_{SL}

Consider an area of dimensions W by W with N ridges higher than h dispersed randomly throughout in both orientation and direction. The lengths of the ridges are such that there is a probability $P_L(L)dL$ of a ridge having a length between L and $L + dL$. We consider that a desired vehicle path begins at the bottom of the area and proceeds north. We neglect lateral offset because of the random orientation of ridges. The probability that any given ridge intersects the vehicle's path is given by:

$$\frac{LP_L(L)dL}{W\pi} |\sin \theta| d\theta$$

where θ ($-\pi/2 \leq \theta \leq \pi/2$) is the angle of the ridge relative to north. For each ridge of length L and orientation θ , the vehicle, on the average, travels an additional distance $L/2 (1 - \cos \theta)$ going around the ridge. Summing over all ridges in the area and over all angles, we have the extra distance ΔD :

$$\begin{aligned} \Delta D &= \frac{N}{2W\pi} \int_0^\infty P_L(L) L^2 dL \int_{-\pi/2}^{\pi/2} (1 - \cos \theta) |\sin \theta| d\theta \\ &= \frac{N \langle L^2 \rangle}{2\pi W} \end{aligned} \quad (4)$$

However, the total number of ridges intersecting the vehicle path is:

$$\mu_h = \frac{N}{W\pi} \int_0^\infty LP_L(L) dL \int_{-\pi/2}^{\pi/2} |\sin \theta| d\theta = \frac{2N \langle L \rangle}{W\pi} \quad (5)$$

Therefore, the total distance traveled over the straight-line distance is

$$\frac{D_T}{D_{SL}} = \frac{W + \Delta D}{W} = \frac{1}{4} \frac{\langle L^2 \rangle}{\langle L \rangle} \mu_h + 1 \quad (6)$$

where μ_h is the number of ridges per unit length with height greater than h encountered along a straight-line path.

The most striking aspect of the analytic result is that the trafficability ratio D_T/D_{SL} depends on the distribution of ridge lengths only through the ratio of the second moment to the first moment. Also, the factor $\langle L^2 \rangle / \langle L \rangle$ only appears linearly, so that small changes in $P_L(L)$ are not extremely critical in determining mobility; this is certainly an advantage since $P_L(L)$ is a difficult distribution function to estimate exactly for sea ice.

HIBLER

Monte Carlo calculation

Clearly the primary defect in the analytic approach is the neglect of ridge intersections. In order to determine the effect of such intersections on trafficability, as well as to verify the analytic model, a Monte Carlo calculation was performed. The calculation proceeded by the following steps.

1. A set of random ridges that would be encountered along a straight-line path was constructed. The number of random ridges having a given orientation θ and length L were weighted by a factor $L \cos \theta$ where θ is the angle between the ridge and the straight-line path. This factor takes into account the increased probability of encountering longer ridges nearly perpendicular to the path of motion. The distribution of ridge lengths was taken to be a negative exponential and was constructed so that $\langle L^2 \rangle / \langle L \rangle = 1.6$ km.

2. Using small increments ΔX , we proceeded along a desired straight-line, searching randomly for a ridge in each increment according to an input ridge frequency μ_h (number of ridges per kilometer above height h).

3. If no ridge was encountered, the desired distance and total distance traveled were increased by ΔX .

4. If a ridge was encountered, the vehicle heading was changed to be parallel to the random ridge of length L at direction θ so that the vehicle still had a forward component of motion.

5. The point at which the ridge was encountered (and thus the distance to go around the ridge) was randomly selected.

6. We then proceeded in ΔX increments at a heading θ , recording after each increment a total distance of ΔX and a desired distance of $(\cos \theta) \Delta X$. In each increment, a random check for other ridges was made.

7. If the end of the ridge was reached before another ridge was encountered, the heading was changed to $\theta=0$ and we returned to step 2 and proceeded as before.

8. If another ridge at direction θ' was encountered while $\theta \neq 0$ (i.e. while we were still going around the first ridge), we declared a cul-de-sac and changed the heading of the vehicle to direction θ' (or $-\theta'$) so that the vehicle had a backward component of motion. We then proceeded as in step 5.

9. The program ended once the total distance D_T reached a desired value. The straight line or desired distance D_{SL} is always less than or equal to D_T .

10. The resulting D_T/D_{SL} ratio and number of cul-de-sacs were recorded as a function of the dimensionless parameter $\mu_h \langle L^2 \rangle / \langle L \rangle$. It is easy to see that D_T/D_{SL} is a unique function of this parameter by dimensional analysis.

One of the more important steps in the Monte Carlo computation is step 8, which takes into account the effects of intersecting ridges. The procedure used does not account exactly for such intersections, but simply declares that the vehicle must have a backward component of motion after encountering a ridge intersection; i.e. it must proceed along a ridge so that it has a negative component of motion in the desired direction. This approximate procedure gives a slightly high estimate of the extra distance traveled but is a reasonable first approximation commensurate with the accuracy of the terrain model. Also, in practice, since the extra distance traveled rises very rapidly as more cul-de-sacs are encountered, a qualitative estimate of the extra distance is adequate for most trafficability purposes.

Note also that, similarly to the analytic result, the trafficability ratio D_T/D_{SL} depends only on $\mu_h \langle L^2 \rangle / \langle L \rangle$, so that uncertainties in the exact moments of the distribution of the ridge lengths may not be critical.

Results of trafficability computations

Random simulations were made using the Monte Carlo computer program described above with a total distance for each simulation of 100 km. For each value of the ridge frequency μ_h (and hence each value of the parameter $\mu_h \langle L^2 \rangle / \langle L \rangle$), six random simulations were performed. For each simulation the trafficability ratio D_T/D_{SL} was recorded, and average results of the six independent simulations were computed for each value of μ_h . The results, as a function of $\mu_h \langle L^2 \rangle / \langle L \rangle$, are shown in Figure 5.

As can be seen from Figure 5, the effect of cul-de-sacs forces the trafficability ratio up rapidly. Since typical values of $\langle L^2 \rangle / \langle L \rangle$ are of the order of 1.5 km, the results indicate that, for more than 2 impassable ridges/km, movement becomes rather tortuous, although possible, up to about 3 ridges/km. The other important

limiting case is the point at which the analytic approximation approaches the Monte Carlo result. From Figure 5, this would appear to be reasonably valid for $\mu_h \langle L^2 \rangle / \langle L \rangle < 1$ or for the presence of less than about 0.5 impassable ridges/km.

Experimental trafficability ratios using overlays

In order to test the trafficability calculations experimentally, ridge overlays of two different regions of the Arctic were prepared using aerial photo mosaics with typical routing paths drawn through them from which trafficability ratios were calculated; the results were compared with predicted results based on the average ridge frequency μ_h of the photo mosaics. The overlays were also used to estimate the nature of the distribution of ridge lengths. The overlays were prepared by using photo mosaics from which dominant ridges were outlined. Although there was considerable approximation, this procedure supplied a qualitative measure of ridges above 5 ft high, especially since shadows might be used to estimate ridge height. Figure 6 illustrates one of the overlays with typical routing paths constructed. In addition to these two photo mosaics, a photo mosaic having a higher ridge density was constructed by adding a set of random ridges to the mosaic in Figure 6 having the same length distribution as the ridges in the figure.

Several routings were constructed on each photo mosaic and average trafficability ratios were computed. A comparison of these experimental trafficability ratios with predicted ratios based on the Monte Carlo calculation and our estimated values of μ_h and $\langle L^2 \rangle / \langle L \rangle$ is given in Table 1. The predicted and observed results are in reasonable agreement, with the observed ratios generally being slightly smaller than the predicted values. Since, as discussed earlier, the Monte Carlo calculation tends to slightly overestimate the trafficability ratio, this type of deviation is not unexpected and might be particularly pronounced when the ridges have a definite lineation, which may have occurred in Figure 6.

Figure 6 also gives a perspective on the amount of maneuvering necessary for a relatively small ridge frequency. For an array of ridges much denser than Figure 6, travel across the region would not be rapid.

Table 1. Comparison of predicted and observed trafficability ratios

Photo mosaic	μ_h	$\langle L^2 \rangle / \langle L \rangle$	$\langle D_T / D_{SL} \rangle_{\text{observed}}$	D_T / D_{SL} Predicted (Monte Carlo)
#1	0.31/km	1.28 km	1.08 ± 0.07	1.10
#2 (Fig.6)	0.65/km	1.54 km	1.16 ± 0.09	1.45
#3 (Constructed)	1.06/km	1.56 km	1.71 ± 0.4	1.80

CONCLUSIONS

The study of sea ice terrain and the formulation of a trafficability model indicate that intersections of pressure ridges make travel over the pack ice by a sizeable (width several meters or larger) vehicle difficult when the vehicle cannot clear most pressure ridges encountered. In particular, if there are more than about 1.0 impassable ridges/km, rapid travel with a sizeable vehicle becomes difficult and, if there are more than 3 impassable ridges/km, any sort of travel with a sizeable vehicle becomes very difficult.

These numbers, in conjunction with information on the distribution of ridge lengths obtained from laser profile data indicate that a vehicle's capability of clearing 5 to 6-ft ridges is a minimal requirement for travel in the Beaufort Sea and central arctic regions, and of clearing 8 to 9-ft ridges is a minimal requirement for travel in the zone off the Canadian Archipelago. More specific trafficability studies may be constructed using the data presented in this paper together with information on the specific nature of the desired mission.

We also note that besides having applicability to sea ice, the calculations in this paper may be applied to any set of linear randomly oriented obstacles.

REFERENCES

1. Weeks, W.F., Kovacs, A. and Hibler, W.D. III, Pressure ridge characteristics in the Arctic coastal environment, Proceedings of the First International Conference on Port and Ocean Engineering under Arctic Conditions, Vol. 1, p 152-183, Department of Port and Ocean Engineering, Technical University of Norway, Trondheim, 1971.
2. Kovacs, A. (1971). On pressured sea ice. Proceedings of the International Sea Ice Conference. National Research Council, Reykjavik, Iceland. pp 276-295.
3. Hibler, W.D. III, and LeSchack, L. (1972). Power Spectrum Analysis of Undersea and Surface Sea Ice Profile, Journal of Glaciology, Vol. II, No. 63, pp 345-356.
4. Hibler, W.D. III (1971) Two Dimensional Statistical Analysis of Arctic Sea Ice Ridges, Proceedings of the International Sea Ice Conference, Reykjavik, Iceland, May 1971, pp 261-275.
5. Hibler, W.D. III, Weeks, W.F. and Mock, S.J. (1972) Statistical Aspects of Sea Ice Ridge Distributions, Journal of Geophysical Research, Vol. 77, No. 30, pp 5954-5970.
6. Mock, S.J., Hartwell, A.D., and Hibler, W.D. III (1972). Spatial aspects of pressure ridge statistics, Journal of Geophysical Research, Vol. 77, No. 30, p 5945-5953.
7. Hibler, W.D. III, Mock, S. J., and Tucker, W.G. III (1974) Classification and Variation of Sea Ice-Ridging in the Western Arctic Basin, Journal of Geophysical Research, in press.
8. Hibler, W.D. III and Ackley, S.F. (1973). Height variation along sea ice pressure ridges and the probability of finding "holes" for vehicle crossings. U.S. Army Cold Regions Research and Engineering Laboratory (USA CRREL) Special Report.
9. Bekker, M.G. (1969) Introduction to Terrain Vehicle Systems. Ann Arbor: University of Michigan Press, p 191-192.



Figure 1. Ground view of "first-year" pressure ridge in the Beaufort Sea, 1971.

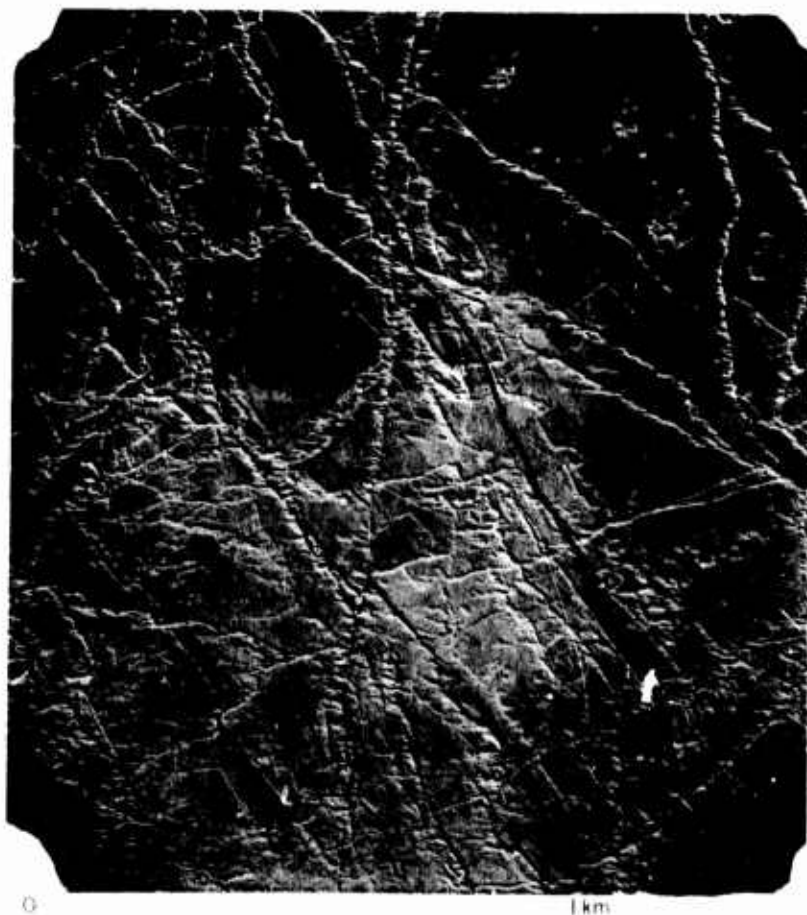


Figure 2. Vertical aerial view of sea ice showing the intersecting nature of ridges. Obtained over the Beaufort Sea by NASA, March 1971.

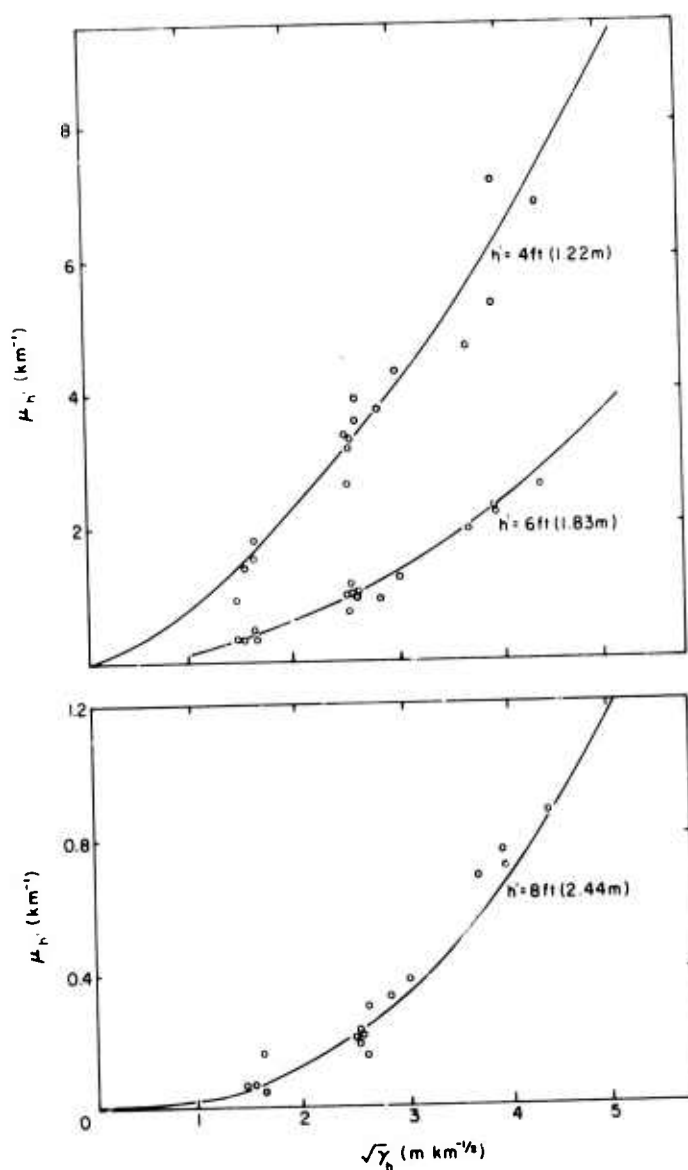


Figure 3. Ridge picker. Predicted (solid lines) and observed values for μ_h , the number of ridges per kilometer above height h' , versus $\sqrt{\gamma_h}$; $h = 4 \text{ ft (1.22 m)}$.

HIBLER

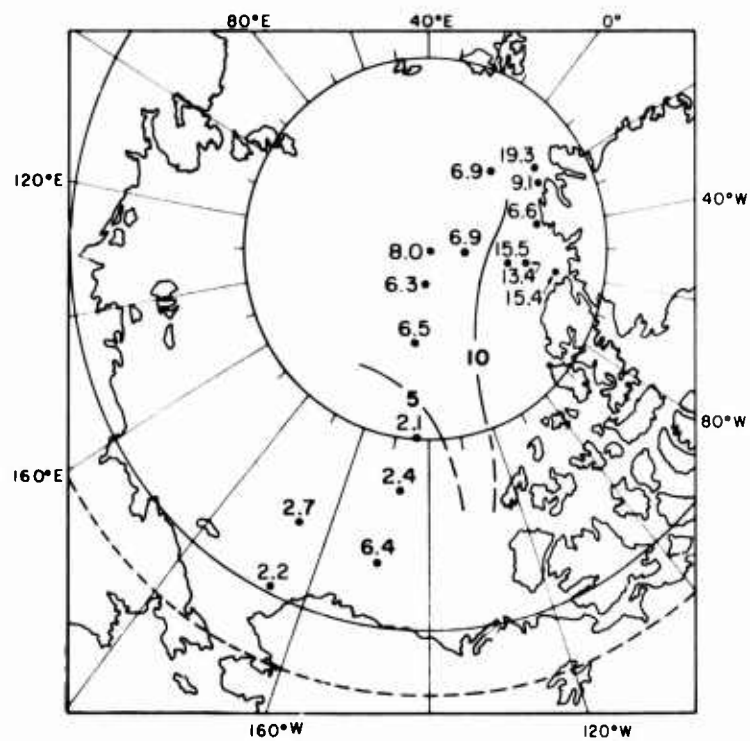


Figure 4. Laser data sample locations and ridging intensities γ_h ($h = 1.22$ m) for November 1970.

HIBLER

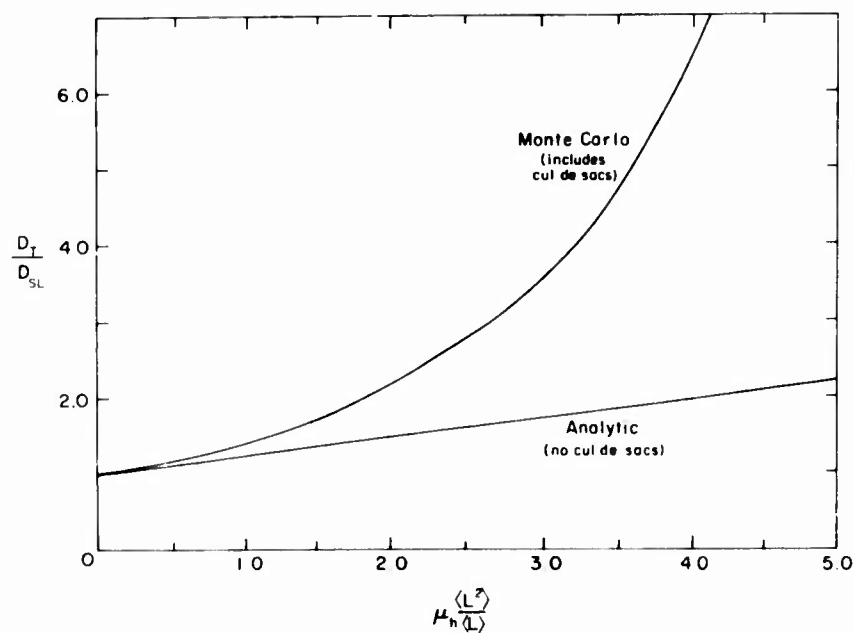


Figure 5. Trafficability ratio D_T/D_{SL} as a function of $\mu_h \langle L^2 \rangle / \langle L \rangle$. μ_h = number of impassable ridges/km, $\langle L \rangle$ = first moment of the ridge length distribution, and $\langle L^2 \rangle$ = second moment of the ridge length distribution.



Figure 6. Ridges prepared from an aerial photo mosaic overlay of sea ice. Simulated vehicle routings are indicated by solid lines with arrows. (Site of March 1971 AIDJEX camp.)

HYDRAULIC TRANSIENT PRESSURES IN PIPING

CHARLES C. HUANG, MR.
*RICHARD J. BRADSHAW, JR., MR.
US ARMY ENGINEER DIVISION, HUNTSVILLE
HUNTSVILLE, ALABAMA

Designers of the SAFEGUARD Ballistic Missile Defense System faced a requirement to design the critical piping systems in protective structures to survive the transient pressures produced in the piping by strong shaking motions resulting from a nuclear detonation. The transient pressure peaks would possibly exceed the design pressures and damage critical equipment. The computer codes available then were for waterhammer analysis and were not suitable for calculating transient pressures resulting from strong shaking motions of the structure. In 1969, the US Army Engineer Division, Huntsville, responsible for the SAFEGUARD facilities design, initiated a R&D program to develop (6) and verify (2) a new computer code for transient pressure calculations for the purpose of designing SAFEGUARD piping systems.

A computer code, HYTRAN (8) was the end product of the development and testing efforts. Eighteen large critical piping systems in the SAFEGUARD System were analyzed by HYTRAN to determine the severity of the transient pressures in the piping and the sizes and locations of accumulators (pressure attenuators) required to protect the critical equipment by lowering the peak pressures and avoiding potential cavitations. The HYTRAN code was validated by laboratory tests.

The purpose of this paper is to introduce HYTRAN to engineers who are concerned with hydraulic transient pressures in fluid filled piping systems subjected to strong shaking motions. The basic formulation, method of solution, cavitation, use of accumulators, modeling techniques, analysis of a typical large system, and comparison of calculated and experimental data are presented.

Preceding page blank

BASIC FORMULATION

Hydraulic transient pressures can be generated in fluid filled piping circuits by a sudden change in flow velocity relative to the pipe. The velocity change may be caused by sudden actuation of valves, fluctuating flow from a pumping source, or a strong shaking of the pipes induced by an earthquake or detonation of weapons. The problem with which this paper is concerned is to examine the transient pressure phenomenon in pipes when strong motions are applied externally to the pipe. In other words, the response of the fluid elements to the base motion of the pipes is the focus of interest. For piping systems that are adequately supported and anchored, the effect of interaction between the oscillating pipes and their contained fluid is usually small and may be ignored.

For one dimensional flow, the equation of motion, Equation 1, and the continuity equation, Equation 2, will provide a complete definition of fluid pressure vs time when the boundary conditions at each end of a pipe are known and the excitation function, $u(t)$ is given.

$$\text{The equation of motion: } \frac{1}{g} \left(\frac{\partial v}{\partial t} + v \frac{\partial v}{\partial x} \right) = -\frac{\partial h}{\partial x} - \frac{f}{2gD} (v|v|) \quad (1)$$

$$\text{The continuity equation: } \frac{\partial H}{\partial t} + v \frac{\partial H}{\partial x} = -\frac{C^2}{g} \frac{\partial v}{\partial x} \quad (2)$$

where $\bar{u} = V - u$; V is the flow velocity; u is the pipe velocity (excitation); \bar{u} is the flow velocity relative to the pipe; f is the friction factor used in Darcy equation for computing flow resistance; H is the total pressure head; and C is the wave velocity, and is expressed by:

$$C = \sqrt{\frac{1}{\frac{f}{g} \left(\frac{1}{B} + \frac{DC_1}{Ee} \right)}} \quad (3)$$

where B is the fluid bulk modulus of elasticity; E is the pipe modulus of elasticity; D is the pipe diameter; e is the pipe wall thickness; and C_1 is a constant ranging from 0.85 to 0.95.

The boundary conditions may be categorized as follows. Three pipes connected by a tee form a three-pipe joint. An elbow or reducer connecting two pipes is designated as a two-pipe joint. The dead end of a pipe is designated as a one-pipe joint as is a supply tank (constant pressure source) or a pump (constant flow source). A typical three-pipe joint (Figure 1a) has three boundaries. The fluid in each pipe may flow across its boundary into or out of the joint. But the conditions of continuity of flow and equilibrium of pressure must be satisfied at the joint. A sign convention which establishes the positive direction of flow becomes necessary in order to account for all possible combinations of the flow patterns at the joint. The sign convention used here is defined as follows: For a horizontal pipe, its right end is designated as its right boundary, R; and its left end as the left boundary, L. For a vertical pipe, L is at its upper end and R at its lower end. The positive direction of flow is assumed to be from L to R. With the convention thus defined, the continuity equations are established by assuming the fluid always leaves a joint across L and enters across R as illustrated in Figure 1. For example, the designations of the joints of the piping circuit shown in Figure 2 can be readily tabulated.

With the boundary designation and the positive direction of flow thus defined, the equilibrium of pressures and continuity of flow at each joint can be correctly expressed. Using the LRR joint in Figure 1b as an example, it is seen that the total head, H , inside the joint must be in balance with those at the boundaries; i.e.,

$$H = H_{P_1} + \frac{K_1}{2g} (\bar{u}_1) |\bar{u}_1| = H_{P_2} - \frac{K_2}{2g} (\bar{u}_2) |\bar{u}_2| = H_{P_3} - \frac{K_3}{2g} (\bar{u}_3) |\bar{u}_3|$$

and that the total flow into the joint must also be in balance with that flowing out; i.e.,

$$-A_1 \bar{u}_1 + A_2 \bar{u}_2 + A_3 \bar{u}_3 = 0$$

where \bar{u}_1 , \bar{u}_2 , and \bar{u}_3 , are the relative flow velocities with respect to the pipe walls and the choice of sign must be consistent with the sign convention already defined.

Now that the boundary conditions of a pipe with one-, two- or three-pipe joints at its ends have been defined, a complete solution of hydraulic transient pressures in a complex, three-dimensional piping circuit becomes feasible.

NUMERICAL SOLUTION

The method of characteristics that has been used for water-hammer analysis (1) is used here for solving the set of equations discussed in the preceding section for each pipe and joint. The HYTRAN code provides the means to do the computation as fully documented in Reference 8.

If we divide the length of a pipe into n small segments of Δx each, there are $n+1$ stations, $x_0, x_1, x_2, \dots, x_n$. The fluid pressure (or flow velocity with respect to the pipe) at each station in an instant of time of t_{i+1} can be determined provided the pressure and velocity at all the stations at t_i are known (See Figure 3). The time step between t_{i+1} and t_i is Δt , which is one of the two basic parameters that must be established prior to numerical computations. The other parameter is Δx , the length of the pipe between two stations, at which the pressure and velocity are to be computed. It is evident that before excitations are applied and transient pressure computation initiated, a steady-state pressure and flow velocity for every pipe of the circuit must be established. Therefore, a steady-state flow analysis precedes the desired transient analysis. There are several techniques that are available for steady-state analysis. The HYTRAN code has a module to do this function and it is a part of the code.

The choice of time step, Δt , and the space interval, Δx , for a computation will affect not only the length of the computation and associated cost, but, more importantly, the resolution of the computed pressures. The nature of the excitation function (the motion applied to the pipes) should be considered in establishing Δt . For example, if a complex waveform excitation containing many frequency components is applied to the pipes, a Δt should be selected such that $\Delta t \leq 0.10T$. T is the period of the highest frequency component. With Δt thus established, the criterion for selecting the value of Δx is that $\Delta x > (C)(\Delta t)$, where C is the wave velocity as expressed in Equation 3. Otherwise, the numerical process of using the method of characteristics will break down and reach a state of instability. In conclusion, the nature of the excitation governs the selection of both Δt and Δx and the total number of computations involved.

COLUMN SEPARATION AND ACCUMULATORS

While the magnitude of the transient pressure in a pipe varies, the local static pressure may drop below the vapor pressure of the liquid; in which case, cavitation may occur. If a cavity thus produced is allowed to grow to the extent that a vapor column is formed, a high pressure generated by the reattachment of the separated fluid column could be damaging. The effectiveness of using accumulators was demonstrated by HYTRAN computation and verified by tests. Figure 4 and Figure 5 indicate the effectiveness of the type of accumulator tested. The computed data are slightly conservative as compared with the test data and provide a margin of safety for design. Therefore, the HYTRAN code is also useful for sizing accumulators.

Fast response and repeatability are the two performance requirements for the accumulators selected for transient pressure attenuation. The type of accumulator adopted for SAFEGUARD application is shown in Figure 6. It consists of an inner perforated metal flow tube having the I.D. of the entrance and exit pipes connected to the unit. On the O.D. of the tube there is an elastomeric diaphragm which separates the fluid from a gas chamber. The chamber is precharged so that the diaphragm assumes an intermediate position under steady-state operating conditions. The diaphragm is free to expand or contract in response to the elasticity of the diaphragm and the flow restriction through the perforated tube.

MODELING TECHNIQUE

A physical piping system must first be idealized by a model that preserves the basic characteristics of the physical system and yet is simple and amenable to efficient computation. Short pipes may be combined with larger ones; identical pieces of equipment that are closely spaced may be combined as one equivalent unit. The flow rate and flow resistance of an equivalent unit must remain the same as the combined flow rate and resistance of the individual units. When modeling a piece of equipment, it is often difficult to obtain the exact configuration of piping circuits contained in the equipment. As a minimum, however, the equipment manufacturer should be requested to provide data on flow rate and pressure drop between the inlet and outlet of the equipment; a general description of the size, diameter, and material of piping; and the maximum tolerable hydraulic pressures determined by proof tests. Using these data, a realistic model can be prepared. A group of pipes

of appropriate lengths and diameters should be selected to match the specific flow rate and pressure drop. The designation of pipes and joints must be made in accordance with the established sign convention to obtain the correct boundary conditions.

Sometimes it is necessary to analyze a system that contains piping of different material. To expedite the computation, it is desirable to use a single value of the pipe modulus of elasticity, E . The procedure for using a single value of E to simulate pipes having different moduli of elasticity requires the consideration of the wave velocity of the contained fluid. As can be seen from Equation 3, the wave velocity, C , depends upon f , g , B , D , C_1 ; E and e . When g , B , D , and C_1 , are system constants and if f is assumed to be constant, the only two remaining variables which affect the value of C are E and e . Therefore, the product of E and e used to calculate the C in the actual pipe must remain unchanged. It follows that to maintain C , the wall thickness of the pipe should be adjusted so that $(Ee)_{\text{actual}} = (Ee)_{\text{equivalent}}$.

Modeling a large physical system without using a large number of pipes is desirable, but over simplification may lead to erroneous results. A parameteric study was conducted to investigate the effect of a set of models representing the same physical system but each having different numbers of pipes and joints. A radar amplifier modulator unit, Figure 7, was instrumented with accelerometers and pressure transducers and tested on a shake table to acquire a set of data. The same unit was first modeled with 49 pipes, and the computed results were compared with the test data as shown in Figure 8, which indicates excellent correlation. The model was then altered to include different numbers of pipes ranging from 59 to 4. The model of 59 pipes represents a realistic model and that of 4 pipes an exceedingly crude model. The pressure-time histories obtained from these models are shown in Figure 9, which indicates that there is only a 20% difference in peak pressure between the 59 pipe model and the 29 pipe model. The peak pressure of the 4 pipe model yields unrealistically high pressures. It may be stated that the results of the computation are not highly sensitive to the number of pipes and joints in a model, but that extremely crude models should be avoided.

ANALYSIS OF LARGE SYSTEMS

One of the eighteen piping systems in the SAFEGUARD System analyzed by HYTRAN is shown in Figure 10. It provides cooling fluid for the radar antenna arrays at the four faces of the turret atop a radar building. The basic design parameters for the components in the system are tabulated in Table 1. The model contains 120 pipes and 118 joints as depicted in Figure 11. All equipment was modeled as two-inch piping with the length of pipes being selected to yield the specified pressure drops at required flow rates. The system was first analyzed without any accumulators and then reanalyzed with accumulators located at various points to reduce excessive high pressures and to avoid cavitations. Table 2 lists the pressures at a selected number of pipes. Without accumulators, the transient peak pressures exceeded the allowable pressure in many places and cavitations also took place. The use of accumulators effectively reduced the excessive transient pressures and eliminated cavitation at selected points.

TABLE I
PIPING CIRCUIT DATA

EQUIPMENT NAME & NO.	MAX. FLOW RATE (GPM)	PRESSURE DROP AT MAX. FLOW RATE (PSIG)	MAX. WORKING PRESSURE (PSIG)	MAX. PROOF PRESSURE (PSIG)
Pump Filter Assembly No. 1204	500	64 (inlet pressure 40 psig, dis- charge pressure 104 psig)	120	200
Heat Exchanger No. 1200	500	17	120	200
Cool 1	150			
3	75			
5	25			
7	250	6	50	75
Cool 2	150			
4	75			
6	25			
8	250	6	50	75

TABLE II
SUMMARY OF MAXIMUM AND MINIMUM TRANSIENT PRESSURES
IN THE CIRCUIT WITH AND WITHOUT ACCUMULATORS

Pipe Number	Pipe Size (in)	ABSOLUTE PRESSURE HEAD (Ft)			
		Without Accumulator		With Accumulator	
		Maximum	Minimum	Maximum	Minimum
1	6	607.5	100.17	280.65	256.78
24	6	378.74	0.5	205.73	123.42
26	2	365.42	0.5	159.63	125.87
34	6	371.09	14.11	205.72	123.41
36	2	384.51	0.5	159.90	125.07
48	6	382.54	52.75	188.25	97.54
50	2	351.57	0.5	199.83	69.23
58	6	336.12	0.5	176.54	108.10
60	2	323.81	3.49	203.66	48.19
72	6	297.15	51.14	180.77	62.67
74	2	323.35	0.5	121.20	90.47
82	6	300.96	21.32	180.75	62.65
84	2	287.96	0.5	122.91	89.97
96	6	349.98	0.5	142.97	75.13
98	2	502.05	0.5	167.78	15.01
106	6	467.87	0.5	148.87	63.31
108	2	395.02	0.5	165.55	34.12
116	8	513.35	0.5	205.48	88.01

CONCLUSIONS:

Based on what has been accomplished by HYTRAN, it may be concluded that HYTRAN is an effective engineering tool for the design of large piping systems to survive the effects of transient pressures, including the selection of accumulators for attenuating pressures and eliminating cavitations.

REFERENCES

1. Streeter, V. L. and Lai, C., "Waterhammer Analysis Including Fluid Friction", Transactions of American Society of Civil Engineers, Vol. 128, Part I, 1963.
2. "Experimental Evaluation of Design Methods for Hardened Piping Systems", IIT Research Institute, Report No. Jt185, September 1970.
3. "Work Statement Specification PAR Prototype Hardness Verification Test Program", Bell Telephone Laboratories, June 1970.
4. Parmakian, J., "Waterhammer Analysis" Prentice Hall, Inc., New York, N. Y., 1955.
5. Lister, M. "The Numerical Solutions of Hyperbolic Partial Differential Equations by the Method of Characteristics", Mathematical Method for Digital Computers, Edited by Ralston and Wilf. John Wiley and Sons, Inc., New York, 1960.
6. "HYTRAN Computer Program User's Manual", the Ralph M. Parsons Company, Document No. SAF-41, November 1970.
7. Li, W. H., and Walsh, J. P., "Pressure Generated by Cavitation in a Pipe", Proceedings of the American Society of Civil Engineers, Vol. 90, No. EM6, December 1964.
8. "Hydraulic Transients Program", U.S. Army Engineer Division, Huntsville, HNDTR-73-9-ED-R, 30 April 1973.
9. DeArmond, R. P., and Rouleau, W. T., "Wave Propagation in Viscous, Compressible Liquid Confined in Elastic Tubes" ASME Paper No. 72-FE-23.
10. Huang, C. C., Bradshaw, R. J., and Yen, H. H., "Piping Design For Hydraulic Transient Pressures", The Shock and Vibration Bulletin, Bulletin 44, 1974.

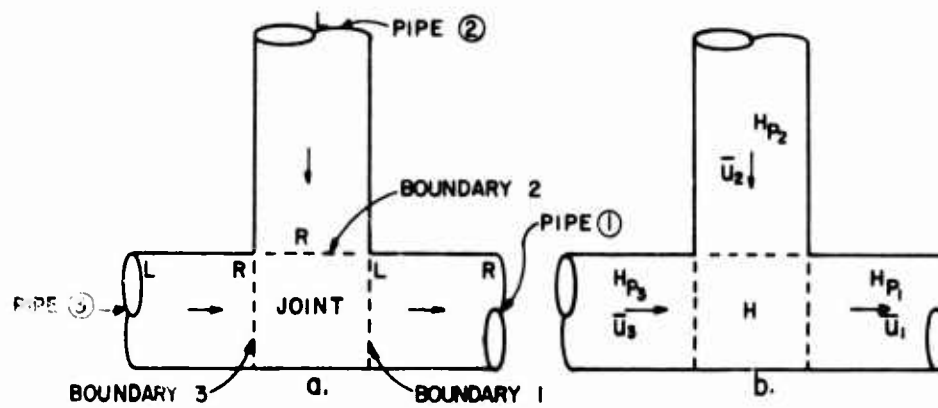


Figure 1. Three-Pipe Joint Boundaries

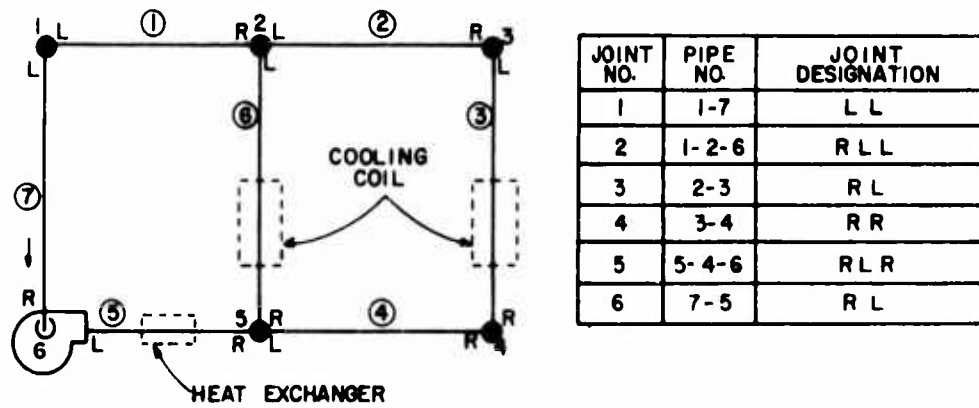


Figure 2. Piping Loop

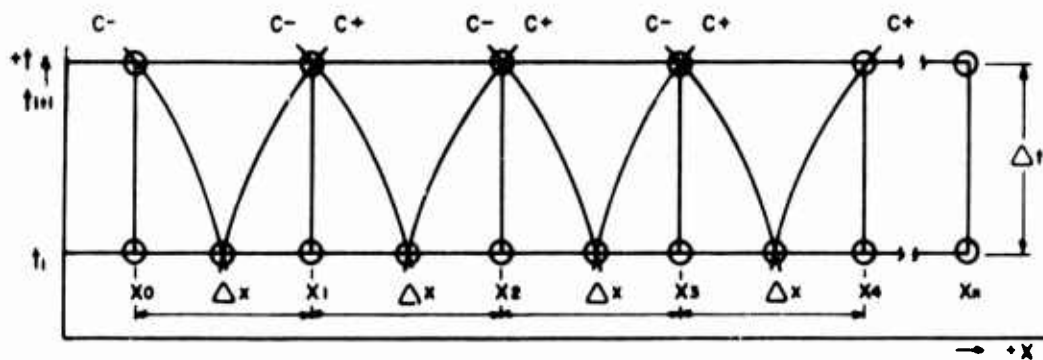


Figure 3. Characteristic Grid

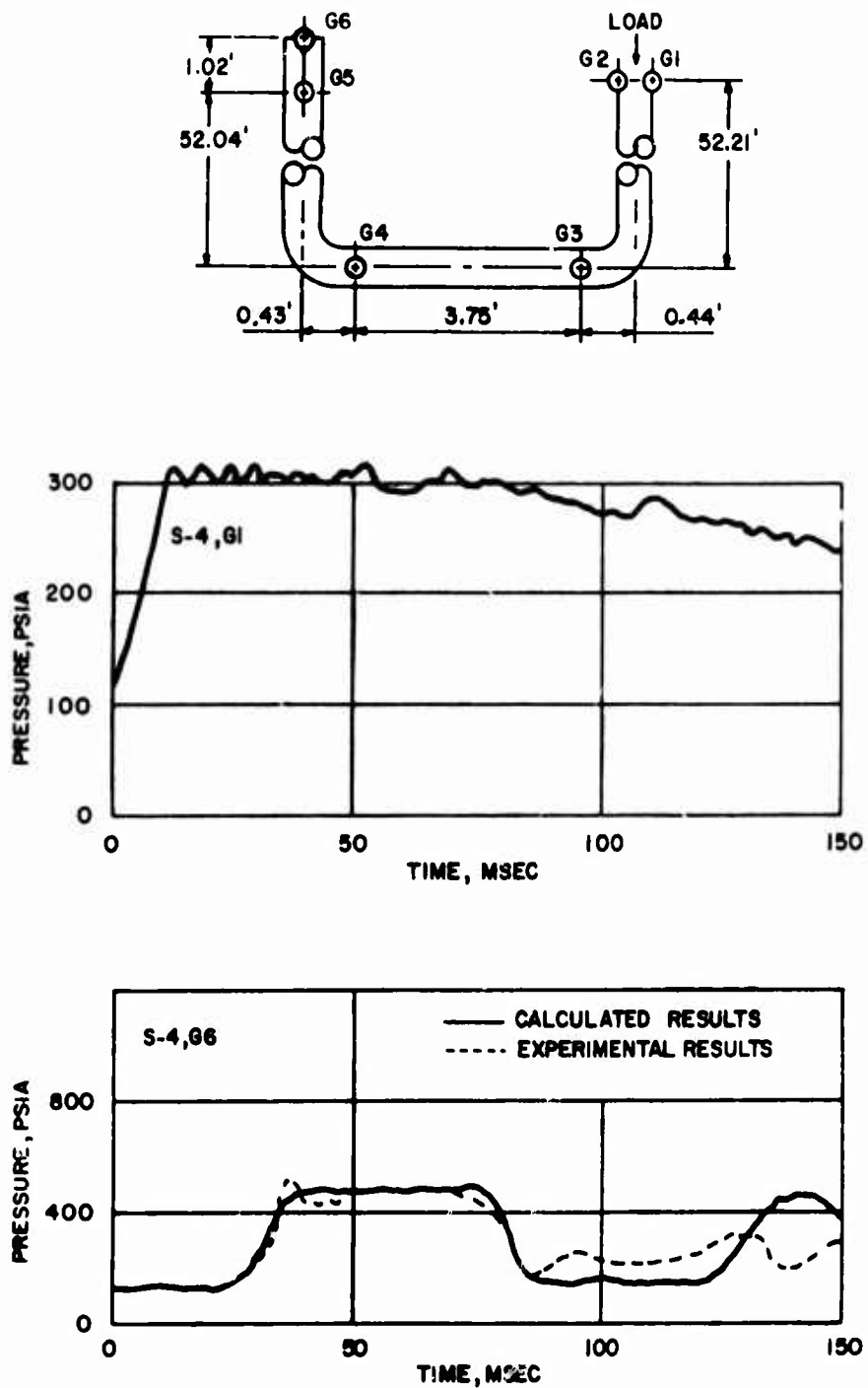


Figure 4. Test Setup Without Accumulator

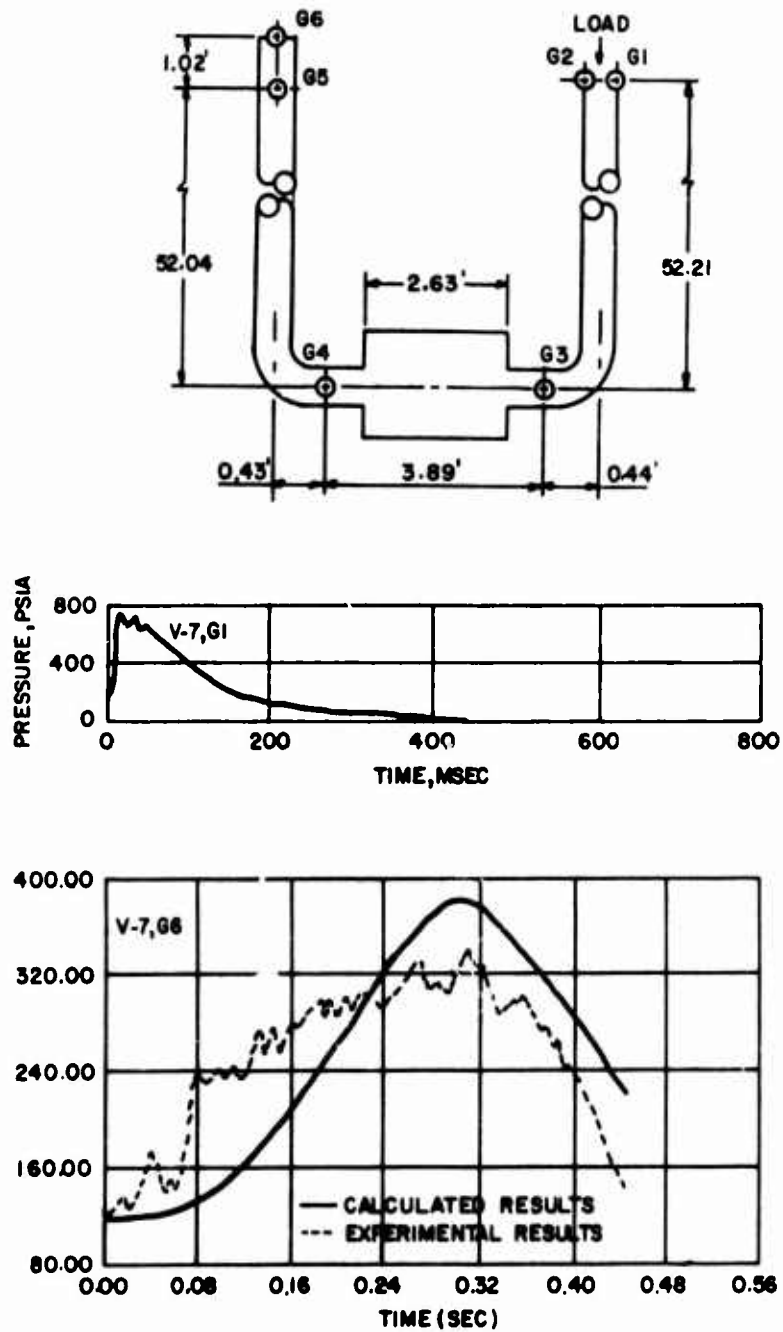


Figure 5. Test Setup With Accumulator

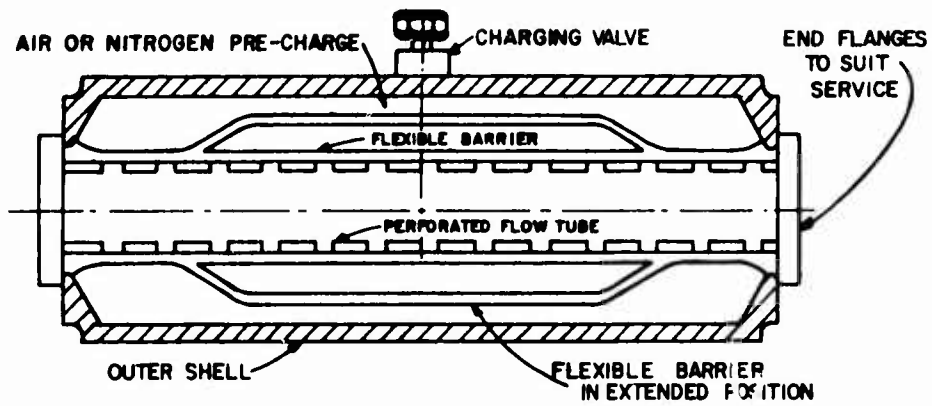


Figure 6. Accumulator

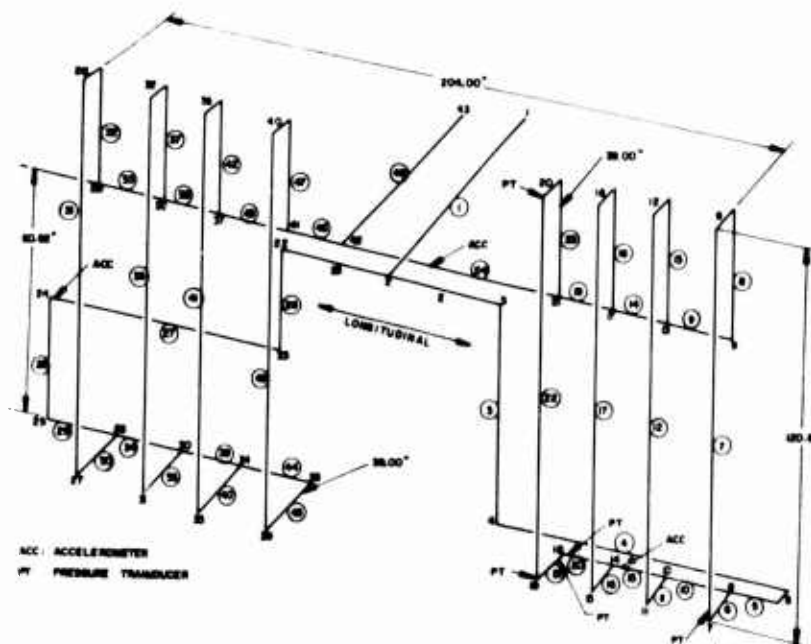


Figure 7. Radar Amplifier Modulator Unit Model

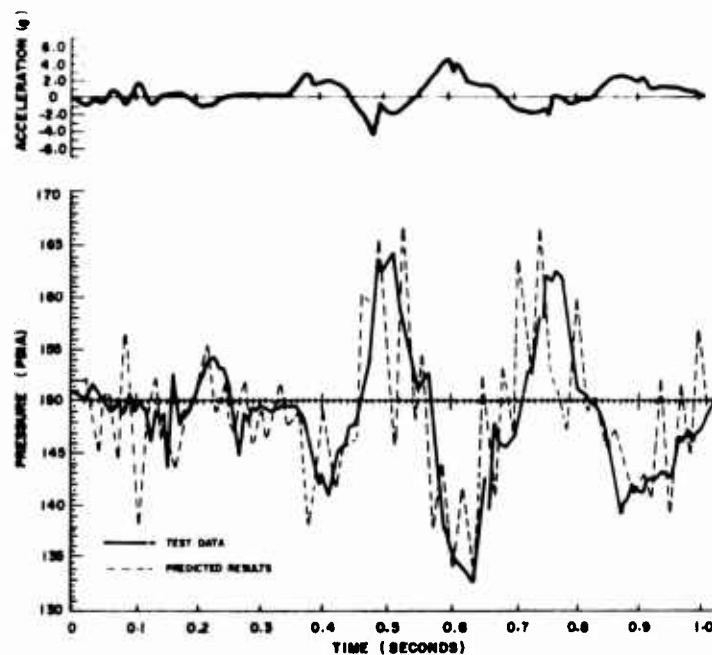


Figure 8. Comparison of Test and Computed Data

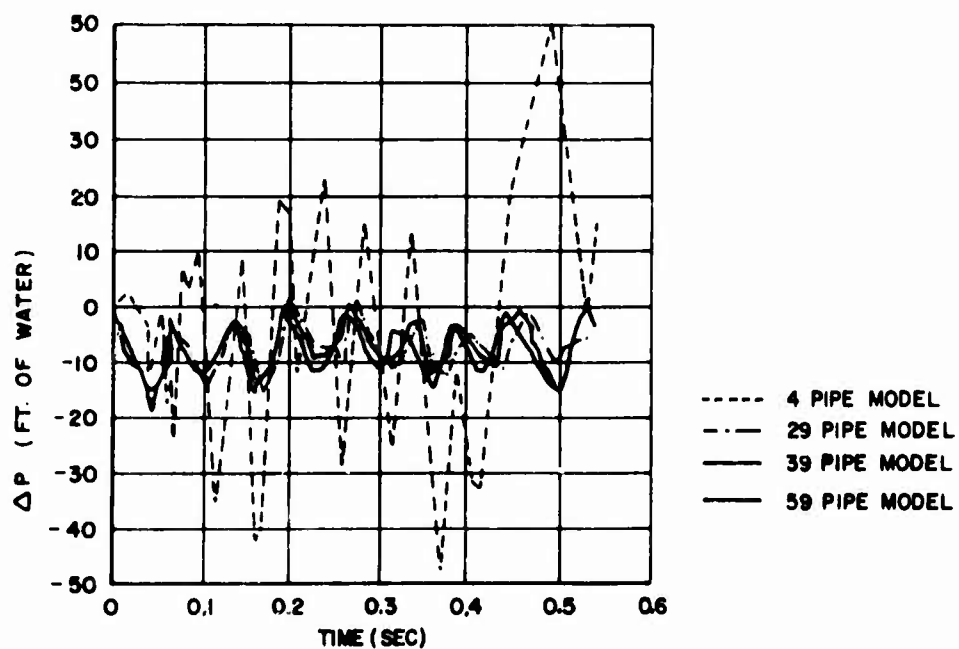


Figure 9. Parametric Study

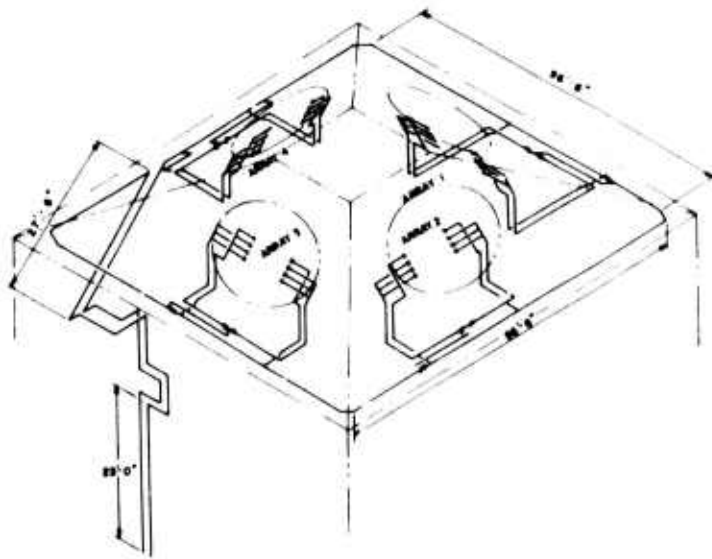


Figure 10. Antenna Array Cooling System

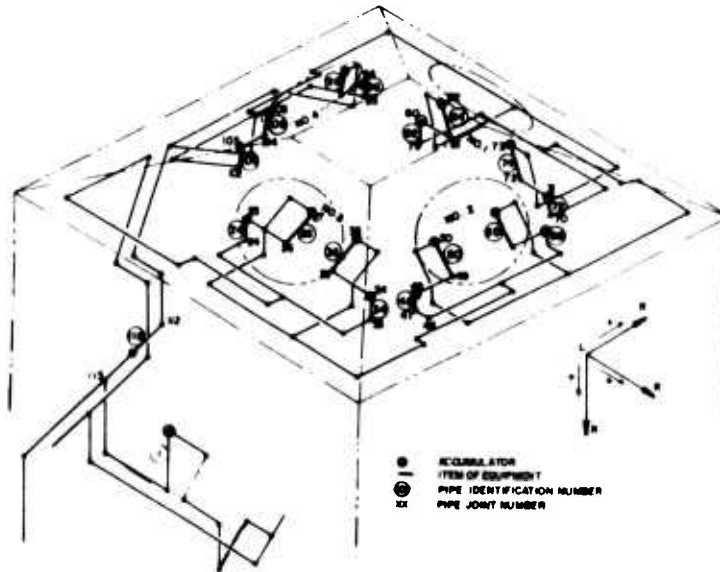


Figure 11. Antenna Array Cooling System Model